

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:10:22 ; Search time 45.5073 Seconds
(without alignments)
4345.419 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730

Sequence: 1 FCLQSTRVRLRNQGHFPST.....MIVKRYSTRSASSQSSR 2057

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	10628	99.0	2058	2 A59267	myosin X - human
2	10011	93.3	2052	2 A59297	myosin X - mouse
3	9861	92.8	2052	2 T18519	myosin X - bovine
4	1921	17.9	2357	2 A59249	class VII unconv
5	1840.5	17.2	2215	2 T30870	myosin VIIa - mous
6	1833	17.1	2175	2 A59255	myosin VIIa, long
7	1716	16.0	2098	2 T25888	hypothetical prote
8	1702.5	15.9	1203	2 A59257	myosin VIIa, short
9	1667	15.8	2121	2 A59233	myosin VII-like pr
10	1593	14.8	3511	2 A59295	unconventional myo
11	1589.5	14.8	3530	2 A59266	unconventional myo
12	1512.5	14.1	1529	2 A59310	unconventional myo
13	1504.5	14.0	1846	2 A59289	myr 6, unconv
14	1500	14.0	2116	2 A26655	myosin heavy chain
15	1492.5	13.9	2022	2 A59256	myosin-IXD [simila
16	1492	13.9	2548	2 E59435	myosin IXA [import
17	1490.5	13.9	2245	2 T18278	myosin heavy chain
18	1487.5	13.9	1611	2 A84743	probable myosin he
19	1487	13.9	2626	2 T31099	myosin-Rhocap prot
20	1466	13.8	1783	2 T42386	unconventional myo
21	1465	13.8	1260	2 T14276	myosin-like protei
22	1463	13.8	1528	2 T14279	myosin-like protei
23	1461	13.7	1980	2 S54307	myosin heavy chain
24	1458.5	13.6	1855	2 A59254	myosin heavy chain
25	1458.5	13.6	1477	2 T00957	myosin heavy chain
26	1458	13.6	1520	2 S46444	myosin MYA1, class
27	1458	13.6	1583	2 T00727	myosin heavy chain
28	1457.5	13.6	1828	2 B59254	myosin heavy chain
29	1450.5	13.5	1556	2 P96587	hypothetical prote

30	1442	13.4	1574	1 A38454	myosin MYO2 - yeast
31	1435	13.4	1515	2 S51824	myosin heavy chain
32	1434.5	13.4	1446	2 T04528	myosin heavy chain
33	1434	13.4	1853	1 A46761	myosin heavy chain
34	1430.5	13.3	1490	2 B4726	probable unconv
35	1429.5	13.3	1830	2 S19188	myosin-V - chicken
36	1424.5	13.3	1242	2 T45976	myosin heavy chain
37	1423	13.3	1374	2 D85390	myosin-like protei
38	1421.5	13.2	1375	2 T05200	myosin heavy chain
39	1399.5	13.0	1502	2 T14278	myosin-like protei
40	1399.5	13.0	1736	2 R6178	hypothetical prote
41	1399	13.0	1516	2 T41235	probable myosin he
42	1394	13.0	1502	2 D84587	probable myosin he
43	1379	12.9	1471	2 T40117	myosin-2 isoform -
44	1373.5	12.8	1509	1 A27224	myosin heavy chain
45	1370.5	12.8	1792	2 T13939	myosin V - fruit f

ALIGNMENTS

RESULT 1

A59267
myosin X - human
C:Species: Homo sapiens (man)
C:Date: 02-Jun-2000 #sequence_reviseion 02-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59267
R:Rogers, M.S.; Streibler, B.E.
Submitted to GenBank, March 2000
A:Description: Identification of myosin X as a specific binding partner for the tumor se
A:Reference number: A59267
A:Accession: A59267
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2058 <ROG>
A:Cross-References: GB:AF234532; P1DN:AA37875.1
A:Genetics:
A:Gene: MYO10
C:Superfamily: myosin motor domain homology
F:66-727/Domain: myosin motor domain homology <MNO>

Query Match	99.0%	Score	10628	DB 2	Length	2058	
Best Local Similarity	99.3%	Pred. No.	0				
Matches	2043	Conservative	5	Mismatches	7	Indels	2
QY	1	FCLQSTRVRLRNQGHFPSTVNSCAEGIVPRRTYGVPTKOSTITOKVTAMHPTEE	60				
DB	4	FTEGTRVRLRNQGHFPSTVNSCAEGIVPRRTYGVPTKOSTITOKVTAMHPTEE	63				
QY	61	GVDDMASLTSLHGGSIMYNLFORYKRNQIMTYIGSILASVNPYQPIAGLYEPATWEOYSR	120				
DB	64	GVDDMASLTSLHGGSIMYNLFORYKRNQIMTYIGSILASVNPYQPIAGLYEPATWEOYSR	123				
QY	121	RHLGELPHRIPALANECYRCILMKRDNOCILIKESGAGKTESTYLILKPLSVISQGLE	180				
DB	124	RHLGELPHRIPALANECYRCILMKRDNOCILIKESGAGKTESTYLILKPLSVISQGLE	183				
QY	191	LSLKEKTSCEVRATLESPIEAFGNATVYNNNSRRGKTVQNLICCKGNIGGRITDC	240				
DB	194	LSLKEKTSCEVRATLESPIEAFGNATVYNNNSRRGKTVQNLICCKGNIGGRITDY	243				
QY	241	ILSSGNRVRONPGERNHIFVALLAGLEHERREBFYSTBENHYLNOQSCVEDKTISD	300				
DB	244	LL-ENKRVYRONPGERNHIFVALLAGLEHERREBFYSTBENHYLNOQSCVEDKTISD	302				
QY	301	QESFEVITAMDVNQFSKEEVREVSRLIAGILHGINETITAGCAQVSFKTALGSAELL	360				
DB	303	QESFEVITAMDVNQFSKEEVREVSRLIAGILHGINETITAGCAQVSFKTALGSAELL	362				
QY	361	GLDPQTLDALTORSMPLRGSEILITPLNVQAAVDSRSLAALVACCEBWTIKKINSRIK	420				
DB	363	GLDPQTLDALTORSMPLRGSEILITPLNVQAAVDSRSLAALVACCEBWTIKKINSRIK	422				

```

QY 421 GNEFKSIGILDI FGFENFENFVNFHFOFNINYNANEKLOEYFNKHIFSLLEOLEYSREGLWME 480
DB 423 GNEDFKSIGILDI FGFENFENFVNFHFOFNINYNANEKLOEYFNKHIFSLLEOLEYSREGLWME 482
QY 481 DIDWINGECGLDIEKKGLGILALINEESHPPQATDSTLBLEKLSQHANNHFYKRVAVN 540
DB 483 DIDWINGECGLDIEKKGLGILALINEESHPPQATDSTLBLEKLSQHANNHFYKRVAVN 542
QY 541 NFGVKYAGEVOYDVAGILEKNDTFRDILNLRSPDFIYDLFEHYSRRNQDTLKC 600
DB 543 NFGVKYAGEVOYDVAGILEKNDTFRDILNLRSPDFIYDLFEHYSRRNQDTLKC 602
QY 601 GSHRRPVTSSQFVK -DSLSHSLMATLSSNPFVRCIKPMNQKPDQFDQAVLNLQLRYSG 660
DB 603 GSHRRPVTSSQFVK -DSLSHSLMATLSSNPFVRCIKPMNQKPDQFDQAVLNLQLRYSG 661
QY 661 MLETVIRKAGYAVRRPFQDYKRYKYLNRNLALPEDVNGKCTSLQLYDASNEQOLK 720
DB 662 MLETVIRKAGYAVRRPFQDYKRYKYLNRNLALPEDVNGKCTSLQLYDASNEQOLK 721
QY 721 TKVFLRESLEQKLEKREESVSHAANVTIRAHVGLFARQYKRYKYLQKNYAPFL 780
DB 722 TKVFLRESLEQKLEKREESVSHAANVTIRAHVGLFARQYKRYKYLQKNYAPFL 781
QY 781 RRRFLHLKKAIVFOQLRGQIARRVYQLLAEKREQEKKQKEEKKRREERERER 840
DB 782 RRRFLHLKKAIVFOQLRGQIARRVYQLLAEKREQEKKQKEEKKRREERERER 841
QY 841 ERREALRAQOEERTKQOELNLOSKOEALITRELEKQENKQYEBELRLKEIEDIQ 900
DB 842 ERREALRAQOEERTKQOELNLOSKOEALITRELEKQENKQYEBELRLKEIEDIQ 901
QY 901 RMKEQOELSLTEASLOKLOERDQELRLSEECRAQOEFLESINDEIDECVRIERSL 960
DB 902 RMKEQOELSLTEASLOKLOERDQELRLSEECRAQOEFLESINDEIDECVRIERSL 961
QY 961 SGGSEFSSLEASACEKKNFNSQPYPEEVDGEFADDAFKDPSNPSHGHSDQRTS 1020
DB 962 SGGSEFSSLEASACEKKNFNSQPYPEEVDGEFADDAFKDPSNPSHGHSDQRTS 1021
QY 1021 GRTSDSSSEEDPYNMDTVPTSPSADSTVILAPSVODSGSLNSSGSESTYCMPONAGD 1080
DB 1022 GRTSDSSSEEDPYNMDTVPTSPSADSTVILAPSVODSGSLNSSGSESTYCMPONAGD 1081
QY 1081 LPSPDGDYDODDYEDGAIITSGSVTFPSNSYQSQSPYRCGCVGYNSSGAYRFSSEGA 1140
DB 1082 LPSPDGDYDODDYEDGAIITSGSVTFPSNSYQSQSPYRCGCVGYNSSGAYRFSSEGA 1141
QY 1141 OSSFEDSEEDPDSRPTDDELSTYRSQVYSCVTLPLYFHSFLYWKGLNNSWKRCWCVKD 1200
DB 1142 OSSFEDSEEDPDSRPTDDELSTYRSQVYSCVTLPLYFHSFLYWKGLNNSWKRCWCVKD 1201
QY 1201 EFTFLMRSHKOEALKQOGLHKKGGSTLSRRMKRWFLRLQSKLMYFENDSEBKLGTV 1260
DB 1202 EFTFLMRSHKOEALKQOGLHKKGGSTLSRRMKRWFLRLQSKLMYFENDSEBKLGTV 1261
QY 1261 EYRTAKEIIDNTTKENGIDIIIMADRTFHLIAESPEDASQMFVSLQVHASTQOEIOBMKD 1320
DB 1262 EYRTAKEIIDNTTKENGIDIIIMADRTFHLIAESPEDASQMFVSLQVHASTQOEIOBMKD 1321
QY 1321 EQANPONAVGLDVLGILDSVCASDSPDRNPSFVIITANVFLHCNADTPEEMHMITLLQR 1380
DB 1322 EQANPONAVGLDVLGILDSVCASDSPDRNPSFVIITANVFLHCNADTPEEMHMITLLQR 1381
QY 1381 SKGDFRVEQOEFTVRMLHKEVNSPKMSLKLKRWFLTLNLSLDYYSSEKNAKLKGT 1440
DB 1382 SKGDFRVEQOEFTVRMLHKEVNSPKMSLKLKRWFLTLNLSLDYYSSEKNAKLKGT 1441
QY 1441 LVNLNSLCSVVPDEKIFKETGYWNVVYGRKCYRLYKTLNLEATWMSVIONVDTTKAP 1500
DB 1442 LVNLNSLCSVVPDEKIFKETGYWNVVYGRKCYRLYKTLNLEATWMSVIONVDTTKAP 1501
QY 1501 IDTPTQOLIQDIKENCLNSDVVEQIYKRNPIILRYTHHPLHSPLLPLPYGINLNLKDKG 1560

```

```

DB 1502 IDTPTQOLIQDIKENCLNSDVVEQIYKRNPIILRYTHHPLHSPLLPLPYGINLNLKDKG 1561
QY 1561 YTTLDDEAIKIFNSIQOLBSMSDPIPIIOGLIQTHDRLPRDELAYCOLIKQTNKVPHG 1620
DB 1562 YTTLDDEAIKIFNSIQOLBSMSDPIPIIOGLIQTHDRLPRDELAYCOLIKQTNKVPHG 1621
QY 1621 SVGNLSWQILTCLSCFPLPSRGLIKYLFHLKRIREQPGTEMEKVALFTYESLKTVC 1680
DB 1622 SVGNLSWQILTCLSCFPLPSRGLIKYLFHLKRIREQPGTEMEKVALFTYESLKTVC 1681
QY 1681 REFVPSRDEIEALHROEKTSTVYCHGGSCKITINSHTTAGEVVEKILRGLAMDSDRM 1740
DB 1682 REFVPSRDEIEALHROEKTSTVYCHGGSCKITINSHTTAGEVVEKILRGLAMDSDRM 1741
QY 1741 FALFEYNGHVDKAIISRTFVADVLAKEFKLATSSEVGLPMKFYKLCFLDITDNPXDS 1800
DB 1742 FALFEYNGHVDKAIISRTFVADVLAKEFKLATSSEVGLPMKFYKLCFLDITDNPXDS 1801
QY 1801 VEFAPMEQAEAVTGHHPAPEENIQLVALRLQYLOGDYTLHAALPPLBEVYSLORLK 1860
DB 1802 VEFAPMEQAEAVTGHHPAPEENIQLVALRLQYLOGDYTLHAALPPLBEVYSLORLK 1861
QY 1861 ARISOSTTFTPCERLEKRTSFLGTLRSPRTGSVVRQKVEEQMDMKEEVSAR 1920
DB 1862 ARISOSTTFTPCERLEKRTSFLGTLRSPRTGSVVRQKVEEQMDMKEEVSAR 1921
QY 1921 ASIIDKMRKPGOMNOEAMAKMALIKEMPGGSLPFVECKEGEPQELMIGVADAVS 1980
DB 1922 ASIIDKMRKPGOMNOEAMAKMALIKEMPGGSLPFVECKEGEPQELMIGVADAVS 1981
QY 1981 VYKRGEGRPLEVFOYEHILSFCAPLANTYKIVDERELLFETSEVVDVAKLMKAYISMIV 2040
DB 1982 VYKRGEGRPLEVFOYEHILSFCAPLANTYKIVDERELLFETSEVVDVAKLMKAYISMIV 2041
QY 2041 KKRISTTSASQSSR 2057
DB 2042 KKRISTTSASQSSR 2058

RESULT 2
A59297
myosin X - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59297
R:Yonezawa, S.; Kimura, A.; Koshiba, S.; Masaki, S.; Ono, T.; Hanai, A.; Kage
Biochem. Biophys. Res. Commun. 271, 526-533, 2000
A:Title: Mouse myosin X: Molecular architecture and tissue expression as revealed by nor
A:Reference number: A59297
A:Accession: A59297
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2062 <YON>
A:Cross-references: GB:A0249706; PIDN:CAB56466.2
A:Experimental source: strain BALB/c; tissue lib testis
C:Genetics:
A:Gene: myo10
C:Superfamily: myosin motor domain homology
F:66-127/Domain: myosin motor domain homology <MMO>

Query Match 93.3%; Score 10011; DB 2; Length 2062;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 125; Conservative 64; Mismatches 64; Indels 10; Gaps 6;

QY 1 FCLQSTRVLRBNGQHPSTVNSCAGIVFRFTDYGVTYKOSTITTKQKTAHPHPTNEE 60
DB 4 FEPBEGARVLRBNGQHPSTVNSCAGIVFRFTDYGVTYKOSTITTKQKTAHPHPTNEE 63
QY 61 GVDMDASLTSLHAGS IMVNLFORRYRNOIMTYIGSIIASVNPYOPILAGLYEPATWEQYSR 120
DB 64 GVDMDASLTSLHAGS IMVNLFORRYRNOIMTYIGSIIASVNPYOPILAGLYEPATWEQYSR 123

```


121 RHGELPHIPIAIANECEYRCIMKRDNOCTILKSGSAGKTESYKILKFLSVISQOSLE 180
 124 CHLGELPHIPIAIANECEYRCIMKRDNOCTILKSGSAGKTESYKILKFLSVISQOTLD 183
 181 LSLKETSCEVERAIILESSPIWEFGNATVNNNSRRECKFVQULNICOKGNIOGGRYDCC 240
 184 LGLQKETSSEVEAIIQSSPIWEFGNATVNNNSRRECKFVQULNICQOGNIOGGRYDCC 243
 241 ILSSONRVNONGPGRNRYHIFYALLAGLEHEERESEFYLPENRYHYNLQSGCEDKTSID 300
 244 LL-EKORVVRIRIPGRNRYHIFYALLAGLEHEERESEFYLPENRYHYNLQSGCEDKTSID 302
 303 QESFEVITAMVQWQFSKEVEEVLRLAGILHLGNIEFTTAGCAQVSPKALGRSADLL 362
 301 QESFEVITAMVQWQFSKEVEEVLRLAGILHLGNIEFTTAGCAQVSPKALGRSADLL 360
 361 GLDPFQULDALTORSMPLRGEIITPLVQOAVDSRDSLAMALVACCEPMVKKIINSRIX 420
 363 GLDPFQULDALTORSMPLRGEIITPLVQOAVDSRDSLAMALVACCEPMVKKIINSRIX 422
 421 GNEBDSIGIIDLIFGFENFVNHPEQFINVYANLQKQEFYNGHIFSLQLEYSRGLVWE 480
 423 GMDKSKSIGIDIFGFENFVNHPEQFINVYANLQKQEFYNGHIFSLQLEYSRGLVWE 482
 481 DIDWIDNEBCLDIEKKLGLLALINEESHFPOATDSTILLEKLSQHANNHFFYKRVAVN 540
 483 DIDWIDNEBCLDIEKKLGLLALINEESHFPOATDSTILLEKLSQHANNHFFYKRVAVN 542
 541 NGVGVHAGEVOYDVRGILEKRDTPRDDLLMLRESRFDITYDLPENYSSRNNDTLKC 600
 543 NGVGVHAGEVOYDVRGILEKRDTPRDDLLMLRESRFDITYDLPENYSSRNNDTLKC 602
 601 GSKHRPPTVSSQPKVDSLSHMAATLSSNPFVNCCKPMQKMPQDPDAVVLNQLRSG 660
 603 GSKHRPPTVSSQPKVDSLSHMAATLSSNPFVNCCKPMQKMPQDPDAVVLNQLRSG 661
 661 MLETVRIRKAGYAVRPFQDYKRYKVLNRNLALPDEVKQCTSLQLYDASNSEMOQLK 720
 663 MLETVRIRKAGYAVRPFQDYKRYKVLNRNLALPDEVKQCTSLQLYDASNSEMOQLK 721
 721 TVVFLRESLEQLEKREBEVSHAAVIRAAVLGIFLAKQYKVLVYIIOKNTAFLL 780
 722 TVVFLRESLEQLEKREBEVSHAAVIRAAVLGIFLAKQYKVLVYIIOKNTAFLL 781
 781 RRRFLHLKKAALVFOKQALGOIARRVYROLALAEKROEKKOBEKKKREBEERER 840
 782 RRRFLHLKKAALVFOKQALGOIARRVYROLALAEKROEKKOBEKKKREBEERER 839
 841 ERREAB-LRAQOEBEETRKOELALQKSQKEALTELEKOKENKQVEIILKEIEDL 899
 840 AGRADLIRAHQEAETRRQELALQKSQRAADITRELEKOKENKQVEIILKEIEDL 899
 900 QMKKEQOELSTLEASIQKQERRDOELRLLEBACPAOEPLESLNFBIDEICVNRNERS 959
 900 QMKKEQOELSTLEASIQKQERRDOELRLLEBACPAOEPLESLNFBIDEICVNRNERS 959
 960 LSGGSFSS-----SELASACCEKPNFNSQYPP-BEEVNEGFEADDDAKDSDPNSEH 1014
 960 LSGGSFSS-----SELASACCEKPNFNSQYPP-BEEVNEGFEADDDAKDSDPNSEH 1019
 1015 SPQRTSGIRTSDDSEEDPYMDTVPTSPSADSVYLLAPSYODSGSLHNSSGESTYCM 1074
 1020 SPQRTSGIRTSDDSEEDPYMDTVPTSPSADSVYLLAPSYODSGSLHNSSGESTYCM 1079
 1075 PGNAGDLPPDGDYDQDDYEDGATITSGSVTFNSYSQNSPDYRGCVGTYNSGAYR 1134
 1080 PGNAGDLPPDGDYDQDDYEDGATITSGSVTFNSYSQNSPDYRGCVGTYNSGAYR 1139
 1135 FSSSEGQSSFFEDSEEDPDRFTDDSLSTRDSVSCVTLPLPHSLYKMGGLMNSMKR 1194
 1140 FSSSEGQSSFFEDSEEDPDRFTDDSLSTRDSVSCVTLPLPHSLYKMGGLMNSMKR 1199
 1195 WCVLXDETFLWFRSKQEBALKQGLHKKGGSSSTLSRRNWKGFVLRQSKLWYFENDSEE 1254

1200 WCVLXDETFLWFRSKQEBALKQGLHKKGGSSSTLSRRNWKGFVLRQSKLWYFENDSEE 1259
 1255 KLKGTVEVTRAKETIINDNTKENGIDITMAADTFLHLAESPDASQWFSVLSQVASTQOE 1314
 1260 KLKGTVEVTRAKETIINDNTKENGIDITMAADTFLHLAESPDASQWFSVLSQVASTQOE 1319
 1315 IOENHDEQANQNVGLTDVGLIDSVCSADSPDRPNSFVITANRVLHCNADTEPEMHW 1374
 1320 IREHNEDEQANQNVGLTDVGLIDSVCSADSPDRPNSFVITANRVLHCNADTEPEMHW 1379
 1375 ITLLQSRSGDTRVSGEYIVNGWHLKEVKNSPKNSSLKLRWFLVLTNSLDYKSSSKN 1434
 1380 ITLLQSRSGDTRVSGEYIVNGWHLKEVKNSPKNSSLKLRWFLVLTNSLDYKSSSKN 1439
 1435 ALKGTVLNLSGCVVPPDEKIFKETGYMNTVYGRKCYRLTYTLLEATRMSSVIONV 1494
 1440 ALKGTVLNLSGCVVPPDEKIFKETGYMNTVYGRKCYRLTYTLLEATRMSSVIONV 1499
 1495 TDTKAPIDTPTQOOLIQDIXENCLNSDVVEQIYKRNPIILRYTHPLHSPLLPYGDINLN 1554
 1500 TDTKAPIDTPTQOOLIQDIXENCLNSDVVEQIYKRNPIILRYTHPLHSPLLPYGDINLN 1559
 1555 LKDKGYTTLODEAIKIFNSIQOLESMSDPIPIIOGIIQOTGHDLRPLDBLYCOLIKOTN 1614
 1560 LKDKGYTTLODEAIKIFNSIQOLESMSDPIPIIOGIIQOTGHDLRPLDBLYCOLIKOTN 1619
 1615 KVPHPGSVGNLYSMOILCLSCFTPLPSRGIILKYVLFKHLKIRBOFGPTEMEKVALFTYES 1674
 1620 KVPHPGSVGNLYSMOILCLSCFTPLPSRGIILKYVLFKHLKIRBOFGPTEMEKVALFTYES 1679
 1675 LKKTCKREFVPSRDEIEMHROENTSVYCHGGSCKITINSHTAGEVVEKLRGLAM 1734
 1680 LKKTCKREFVPSRDEIEMHROENTSVYCHGGSCKITINSHTAGEVVEKLRGLAM 1739
 1735 EDSRNMFLPEYNGVADVAIESRTVAVDLAKFEKLAATSVGDLPMKFFYKLCFLDITD 1794
 1740 EDSRNMFLPEYNGVADVAIESRTVAVDLAKFEKLAATSVGDLPMKFFYKLCFLDITD 1799
 1795 NVPKDSVFAPFEPBAHVAIVHGHHPAPEBNLOVLAALRYLOQDYTLHAIPLPLEEY 1854
 1800 NVPKDSVFAPFEPBAHVAIVHGHHPAPEBNLOVLAALRYLOQDYTLHAIPLPLEEY 1859
 1855 SLQRLKARISOSTKFTPCERLEKRTSFLBGLTARSPTS SVVQKYEBOQLDMWIK 1914
 1860 SVQRLKARISOSTKFTPCERLEKRTSFLBGLTARSPTS SVVQKYEBOQLDMWIK 1919
 1915 EVSSARASIIDKRRKFGQANQOQAMAKYMALIKEMPGYSTLFDVECKEGGFPQELMIGV 1974
 1920 EVSSARASIIDKRRKFGQANQOQAMAKYMALIKEMPGYSTLFDVECKEGGFPQELMIGV 1979
 1975 SADAVSVKRGREGRLVFOYEHLISFGAPLANTYKIVYVDERELLFETSEVVDVAKLMA 2034
 1980 SAEAVSVKRGREGRLVFOYEHLISFGAPLANTYKIVYVDERELLFETSEVVDVAKLMA 2039
 2035 YISMVKKRYSTTRSSASQSSR 2057
 2040 YISMVKKRYSTTRSSASQSSR 2062

RESULT 3
 T18519
 myosin X - bovine
 C/Species: Bos primigenius taurus (cattle)
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Feb-2001
 C/Accession: T18519
 R/ Corey, D.P.; Dertler, B.H.; Solc, C.K.; Duyk, G.M.; Cheney, R.B.
 submitted to the EMBL Data Library, April 1996
 A/Description: Cloning and expression of myosin X, a novel unconventional myosin with pl
 A/Reference number: Z18942
 A/Accession: T18519
 A/Status: preliminary/ translated from GB/EMBL/DBJ
 A/molecule type: mRNA

A/Residues: 1-2052 <COR>
A/Cross-references: EMBL:U55042; NID:G1755048; PID:G1755049; PIDN:AAB39486.1
A/Experimental source: aorta
C/Superfamily: myosin motor domain homology
C/Keywords: nucleotide binding; P-loop
F/66-727/Domain: myosin motor domain homology <MOC>
F/157-164/Region: nucleotide-binding motif A (P-loop)

Query Match 92.8%; Score 9961; DB 2; Length 2052;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 1909; Conservative 76; Mismatches 62; Indels 12; Gaps 6;

```
QY 1 FCLQGRVWLRNGQHFPSTVNSCAGIIVFRTDYQVFTYKOSTITTHQKVTAMAPTNEE 60
DB 4 FPEEGRWVLRNGQHFPSTVNSCAGIIVFRTDYQVFTYKOSTITTHQKVTAMAPTNEE 63
QY 61 GVDMMATLTELHGSAIMNLVQRYKKNQIYTYIGSIASVNPYKTIYTGYSRDAVDRYER 120
DB 64 GVDMMATLTELHGSAIMNLVQRYKKNQIYTYIGSIASVNPYKTIYTGYSRDAVDRYER 123
QY 121 RHUGELPHIIFAIANECYRCIMGRHNOCLIKGESGAGKTESTKILKPLSVISQOSLE 180
DB 124 CHGELPPIHFAIANECYRCIMGRHNOCLIKGESGAGKTESTKILKPLSVISQOSLD 183
QY 181 LSIKERTSCVERAILESSPIMEAFGNAKTVYNNSSRFKGVQNLICQKGNIQGRIYDC 240
DB 184 LSIKERTSCVERAILESSPIMEAFGNAKTVYNNSSRFKGVQNLICQKGNIQGRIYDY 243
QY 241 ILSORRVVRONGENNYHIFYALLAGLHEEBEETVLTSPREYHNLNLSGCEDETTID 300
DB 244 LL-EKRRVVRONGENNYHIFYALLAGLHEEBEETVLTSPREYHNLNLSGCEDETTID 302
QY 301 QESFREVTAMDMQFSKEKEVREVSRLAGILHIGNIEFTAGAGVSEFKALGRSAELL 360
DB 303 QESFREVTAMDMQFSKEKEVREVSRLAGILHIGNIEFTAGAGVSEFKALGRSAELL 362
QY 361 GLDPTQLTALTORSHFLRGEELITPLNVOQAVDSRSLAMALYACCFEMVIKIKINSRIK 420
DB 363 GLDPTQLTALTORSHFLRGEELITPLNVOQAVDSRSLAMALYACCFEMVIKIKINSRIK 422
QY 421 GNEDEFSIGLIDFGFENFVNHFEQPINYANERKIOEYFNKGIIFSELBOLEYREBLVWE 480
DB 423 GNEDEFSIGLIDFGFENFVNHFEQPINYANERKIOEYFNKGIIFSELBOLEYREBLVWE 482
QY 481 DIDMDINGECLDLIEKKGLLALINESHPQATDSTLEKLSQANHHFYKPRVAVN 540
DB 483 DIDMDINGECLDLIEKKGLLALINESHPQATDSTLEKLSQANHHFYKPRVAVN 542
QY 541 NFGVKYAGEVOYDVAGILEKNRDPFRDULMLRESRPFYIDLEFHVSSRNNQDTLKC 600
DB 543 NFGVKYAGEVOYDVAGILEKNRDPFRDULMLRESRPFYIDLEFHVSSRNNQDTLKC 602
QY 601 GSKHRRPTVSSQFKVDSLHSLMATTSSNPFPRCCKPMQKMPDFDQAVVNNQIRYSG 661
DB 603 GSKHRRPTVSSQFKVDSLHSLMATTSSNPFPRCCKPMQKMPDFDQAVVNNQIRYSG 661
QY 661 MEETVIRKAGAVARRPQDFYKRYKVLNRNLAPEDVNGKCTSLQLDASNSEWQOLK 720
DB 662 MEETVIRKAGAVARRPQDFYKRYKVLNRNLAPEDVNGKCTSLQLDASNSEWQOLK 721
QY 721 TYVFLRESLEQLEKKREEVSHAMAVIRAHVGLFARQKRYKVLVIGQNYAFALL 780
DB 722 TYVFLRESLEQLEKKREEVSHAMAVIRAHVGLFARQKRYKVLVIGQNYAFALL 781
QY 781 RRRFLHLKKAATVFOQLAGQIARVYRQLAEKREBEKKQREBEKKREBEERER 840
DB 782 RRRFLHLKKAATVFOQLAGQIARVYRQLAEKREBEKKQREBEKKREBEERER 841
QY 841 ERREARLRAQOEETKQOELAL--QSKQKALITRELEKOKENKQVEIILLEKEIEDL 899
DB 842 ERREARLRAQOEETKQOELAL--QSKQKALITRELEKOKENKQVEIILLEKEIEDL 901
QY 900 QRMKEQOELSTLEASTLQQLQERRDDELRLLEBAACRAAQEFLESINFDIEDECVRNIERS 959
```

```
DB 902 QRMKEQOELSTLEASTLQQLQERRDDELRLLEBAACRAAQEFLESINFDIEDECVRNIERS 961
QY 960 LSGSEFSELEAESCEKPNPNFNSQPP--BEVVEGEGRADDAKDSDPNSEHSHDOR 1018
DB 962 LSVGSGCTGE--QGAKEKPFNSQPP--BEVVEGEGRADDAKDSDPNSEHSHDOR 1019
QY 1019 TSGIRTSDDSEEDPYMNDTVTPSPSADSTVLLAPVQDSGSLHNSSGSESTYQMPQNA 1078
DB 1020 TSGIRTSDDSEEDPYMNDTVTPSPSADSTVLLAPSE-----HDSAGERTYCLPQRP 1073
QY 1079 GDLPSPDGDDYDQDDYEDGATISGSVTFNSYSQNSPDYRCGVGTNYSSGAYRFSSE 1138
DB 1074 GALPAPBGDYDDQDDYEDGATISGSVTFNSYSQNSPDYRCGVGTNYSSGAYRFSSE 1133
QY 1139 GAQSSFEDESEDPDRPTDDELSTRSDSVYSCVLPYFHSFLYMKGLMNSMKRKCVL 1198
DB 1134 GAQSSFEDESEDPDRPTDDELSTRSDSVYSCVLPYFHSFLYMKGLMNSMKRKCVL 1193
QY 1199 KDETFLMFRSKQKALKQMLHKKGSSSTLSRNNKCKMFVLROSKLMTYFENDSEBKLG 1258
DB 1194 KDETFLMFRSKQKALKQMLHKKGSSSTLSRNNKCKMFVLROSKLMTYFENDSEBKLG 1253
QY 1259 TVEVFTAKIINDTYKENGIDITMADRTPHLIAESPEDASQMFSLVSOVHASTDOEIOBM 1318
DB 1254 TVEVFTAKIINDTYKENGIDITMADRTPHLIAESPEDASQMFSLVSOVHASTDOEIREM 1313
QY 1319 HDEQANPNQAVGTLVGLIDSVCAQSDSPDRPSFYTITANVYLHONACTPERMHHMITLL 1378
DB 1314 HDEQANPNQAVGTLVGLIDSVCAQSDSPDRPSFYTITANVYLHONACTPERMHHMITLL 1373
QY 1379 QRSKQDTEVEGQEFIVRGMLHKEVNSPMSLSKLKCKMFVLTHNSLDYKSSSEKVALKL 1438
DB 1374 QRSKQDTEVEGQEFIVRGMLHKEVNSPMSLSKLKCKMFVLTHNSLDYKSSSEKVALKL 1433
QY 1439 GTVLNLSLCVPPPEPKIFKETGYNVTVYGRKHCYRLYTKLNEATMSSYIQNVDTDK 1498
DB 1434 GTVLNLSLCVPPPEPKIFKETGYNVTVYGRKHCYRLYTKLNEATMSSYIQNVDTDK 1493
QY 1499 APIDPPTQQLIDIKENCLNSVVEQIYKRNPLRTHNPHLSPLPLPYGGINLNLKD 1558
DB 1494 APIDPPTQQLIDIKENCLNSVVEQIYKRNPLRTHNPHLSPLPLPYGGINLNLKD 1553
QY 1559 KGYTLLODEALKIFPMSLOQLSMDSPRPIIOGILQTHGDLRLRDELVCOLIKQTNKVP 1618
DB 1554 KGYTLLODEALKIFPMSLOQLSMDSPRPIIOGILQTHGDLRLRDELVCOLIKQTNKVP 1613
QY 1619 PGSVGNLSWQITLCTPLPSRGLKYLKFLHRLRIEOPGTEMEKVALFIYESLTKT 1678
DB 1614 PGSVGNLSWQITLCTPLPSRGLKYLKFLHRLRIEOPGTEMEKVALFIYESLTKT 1673
QY 1679 KCRFVPRDELEALIHQEMTSYVCHGGSSCKTINSHTTAGVVEKLRGLMEDSR 1738
DB 1674 KCRFVPRDELEALIHQEMTSYVCHGGSSCKTINSHTTAGVVEKLRGLMEDSR 1733
QY 1739 NMFLALFEYNGHVDAKIESRTVADVLAKEFKLAATSEGDGDPWKYFYLQCYFLDNDVPK 1798
DB 1734 NMFLALFEYNGHVDAKIESRTVADVLAKEFKLAATSEVGEQPMKRYFYLQCYFLDNDVPK 1793
QY 1799 DSVFPAFMFEQAHAEAVIHGHHPAPEBNIOVLAALRYLOQDGYAPHAIVPPLVEYYSQOR 1858
DB 1794 DSVFPAFMFEQAHAEAVIHGHHPAPEBNIOVLAALRYLOQDGYAPHAIVPPLVEYYSQOR 1853
QY 1859 LKARISOSTKFTPEERLEKRTSLBGTLLRSRPTGSVVAPQKVEEOMLMWIKEEVS 1918
DB 1854 LKARISOSTKFTPEERLEKRTSLBGTLLRSRPTGSAIQAKEEBQVMDVMVEBCS 1913
QY 1919 ARAAIIIDMKRFQGNQOQAMAKYALIKEMPGYSTLFVDECKEGGFPQOELMIGVSADA 1978
DB 1914 ARAAIIIDMKRFQGNQOQAMAKYALIKEMPGYSTLFVDECKEGGFPQOELMIGVSADA 1973
QY 1979 VSVYRGGGRPLEVFOYHHIISFGAPLANTYKIVYDERELLFETSEVVDVAKMKAYISM 2038
```



```

Db      1532 CEHQSSESLPYKKNF-----VLAALNNGINIVDPATSKMLESYKVN-----CQSQ 1575
Qy      1478 TKLNATATWSSVIONVDTTKAPIDPTQOLIODIDENCLN-----SVV 1521
Db      1576 QNLKSDKSVSILLENKSTLQA-PTGDVQKLVSLIKEVSLYLRNNAKYARALKDVAVSDT 1634
Qy      1522 VEOIYKRNPIRLRYT----- 1535
Db      1635 SLLEPKRNDITITTFDQENKMFQGLNGKESFPVDHVEILLSDVPPQPVHVPATLSR 1694
Qy      1536 ----- 1535
Db      1695 PMSPTIPNTNPPPPPSISDSMSPPPOVGMLEPPPPPSVMSGTPKEIPISLGIPPPPS 1754
Qy      1536 -----HHPLEPL--LPLPYGDINLNLKDKGYTT----- 1563
Db      1755 SSNSVSPNSPIGSPMMGIPPPPTISVHSLNSGNSTPPPPLSLSTPPPLSTPPPISSP 1814
Qy      1564 ----- 1563
Db      1815 PNRSSLRVSMNTSNDGNSDDSKLTVSPALIGDSQLAQWASTFRSPKRASTLN 1874
Qy      1564 -----LQDEAIKIPNSLQOLESMSD-PIR- 1586
Db      1875 QOQATLKRKAPVDPTAFYFNKDPKIKESLIEMAKLSKAIKVFSEIMMW--MGDYPIPK 1932
Qy      1587 -----IIIGILDTGHDRLRLDELXCOLIKOTNKVPHSGVGNLWSMOILCTGCTPLP 1640
Db      1933 GGTASLVIGSIIISRGLENHRLDELYICQAVRQTNKPKVESAKK--GFELIIFLSTTSP 1990
Qy      1641 SRGILK-YLKFLHKR--IREOPG-----TEMEXALPTYESLTKTKCREPVPBR 1687
Db      1991 SDELDPFMQLSRNIAIOSSPOLASLAVCIKLESLHSPISYQO-----RKMGPSA 2044
Qy      1688 DEELALIHQEMSTYVCHGGGCKT-----TINSHTAGVVEKLRGLAM-ED 1736
Db      2045 TEL-----QSFNSNLENGDISTCKIRFIDOSTKLAKINTYTTIREITPTVCROYGISQO 2098
Qy      1737 SRMFPALFEYN--GHVDKAIESRTVADVLAKEKLAAATSEVDDLPMKFYFLKCYGLDTP 1794
Db      2099 STMFEGISAVNETAGISKVSEETDMLYDVARNEQ--SBEKGE--RYFQVRRRFPID 2151
Qy      1795 NVPK-----DSVEFAFMEOQAEVAVIHG--HHAPAEENLVLAALRLQYLQGYTL 1843
Db      2152 DVVKILDOEHLMTDDDICELTYCQJRDENMGKLYTNVAKOSIIAAILDQLLYNQS- 2210
Qy      1844 HAAIPPLEEYVYSLQRLKARISOSTKTFPCERLEKARTSEFLBGLTRRSFRTGSVVRQVY 1903
Db      2211 -----KLVLTKEVNRQVLP 2224
Qy      1904 EEOMLDMWIKKEVYSARASIIDKMRKFOGNNQOAMAKYM--ALIKEMPGVGSITPDVE 1960
Db      2225 DQILNSQNTKWMISMESQI-----FELVSQTPBYLKLMPFNLJLSKSPDLFGCTIFNQI 2278
Qy      1961 CKEGGEPOEIMLVGSADAVSVY--KRGEGRPLEVFOYEHLISFGALANTYKIVLDER-- 2016
Db      2279 QKEN--PRAMLAINKKGVSIIPDHTKESKNFWTFQGISNVAF--TDITFCIMTGNLAK 2333
Qy      2017 --ELLETSEVDVAVAKMKAYIS 2037
Db      2334 PIKQFTTDEHSSIASVQFYSS 2356

```

RESULT 5

T30870
 myosin VIIa - mouse
 N:Alternate names: Gene shaker-1 protein
 C:Species: Mus musculus (house mouse)
 C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 02-Feb-2001
 C:Accession: T30870, S60346
 R:Walsh, J.; Mburu, P.; Gibson, F.; Liu, X.; Saw, D.; Steel, K.P.; Brown, S.D.M.
 submitted to the EMBL Data Library, December 1996

```

A:Description: Mutation analysis of the mouse myosin VIIa deafness gene.
A:Reference number: 220913
A:Accession: T30870
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-2215 <MAL>
A:Cross-references: EMBL:U81453; NID:g1778381; PID:g1778382; PIDN:AAB40708.1
R:Gibson, F.; Walsh, J.; Mburu, P.; Varela, A.; Brown, K.A.; Antonio, M.; Beisel, K.W.;
Nature 374, 62-64, 1995
A:Title: A type VII myosin encoded by the mouse deafness gene shaker-1.
A:Reference number: S60346; MUID:95174881; PMID:7870172
A:Accession: S60346
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra
A:Molecule type: mRNA
A:Residues: 96-156, 'R', 158-237, 'E', 239-681, 'V', 683-726 <GLB>
A:Note: this sequence variant is from the mouse deafness gene shaker-1.
A:Gene: sh1, MyoVIIa, MGI:104501
A:Cross-references: MGI:104501
A:Map position: 7:48.1
C:Function:
A:Description: seems to be involved in auditory transduction
C:Superfamily: myosin motor domain homology
C:Keywords: nucleotide binding; P-loop
F:68-729/Domain: myosin motor domain homology <MMO>
F:96-726/Domain: myosin motor domain homology (fragment) <MMOT>
F:158-165/Region: nucleotide-binding motif A (P-loop)
Query Match 17.2%; Score 1840.5; DB 2; Length 2215;
Best Local Similarity 26.3%; Pred. No. 2.7e-75;
Matches 637; Conservative 384; Mismatches 817; Indels 585; Gaps 88;
Qy      4 OSTRVWL-RENGQHEPSTVNSCAEGIVFRFDYGVQVFFYKOSTIRH-----QKVTAMH 55
Db      6 KCDYVMMLKSGQEDVPI-----GAVYKLCDSQGIQVVDDEDNHMTSPQATIKIKRH 60
Qy      56 PTNEGVDDMASLTLEHGSGIMVNI-PQYKRNQITVYIGSIIASVNPQIAGLYEPATM 115
Db      61 PTVSHGVEDMIRLQDLNAGILRNLIYRDLIYVGSILVAVNPQ-LLSISPEHI 119
Qy      116 EQYSRHLGELPRHIFAIANEBCYRCLMRHNOCLLIGESGAGSTESTKILKEKLSVS 175
Db      120 RQYTNKKIGEMPHIFALADNCFYMKRNNDQCIIIGESGAGSTESTKILQFLAALS 179
Qy      176 QOSELSLKEKTSCEYRAILLESSPIMEAFNAKATVYNNNSRFKGFVOLNIOGKNIQGG 235
Db      180 GGH-----SWIEQVLENTPILEAFNAKATIRNDSSRFKGIIDHFKRAIEGA 230
Qy      236 RIVDCILSSQNRVVRONGERNYHIFVALLAGLEHEEBEFLSTPENYHYLNSGCVED 295
Db      221 KIEQYLL-SKSRVCRQADDERHYHVFYCMLEGMNEBEKKKLGIGQAADYNYLAMGNCITC 289
Qy      236 KTISQESFREYITMDVMQSKBEVRSYLLAGILFNGNE-----FITAGANQVSPK 350
Db      230 EGRVDSQEVANIRSMKVMKLMFTDENWEISKLLAAILMNGVQYARTPENLDACEVLS 349
Qy      351 TALGRSAEILGLDPQLTDAITORSMPFGEIITPLNVOQAVDSRDSLAMALACCPEM 410
Db      350 PSLATRAELVENPFDLMSCLTSKRLITRGREVSTPSLREGQALDVRAFPVGIYGRLEFM 409
Qy      411 VIKKINSRI-----KGNEDFKSIGIIDI PGFENEVNHFPQFINVANEKLOEYFNKHI 464
Db      410 IVEKINAAIYKPPPLEVKNRSRSIIGLIDIFGFENFTVNSFEQICINFANHEHLQCFVRHV 469
Qy      465 FSLQELYSRGLVVEDIDWDINGCCLLI-EKLGILLALINEESHFPQADSTLLEKTH 523
Db      470 FTLQEEYDLESDILHIEFTDNOALDMINRPNVNSLIDESKFPKGTATYLAHLKLN 529
Qy      524 SOHANHHYVYKPRVA-VNPFGVKHYAGEVQDVRRGILTEKNRDTFFDDLLNLLRESRPFPI 582
Db      530 SOHKLNANYVPPKNSHETQFGINHPAGVYVYESQFLKRNKDTLHGDIILQVHSSRNKFV 589
Qy      583 YDLFEHVSRRNQDTLKGSKRHR--PTVSSQFKVDSLHSLMATLSSNPFFVRCIKPMN 640

```

```

Db      590 KOIF-----QADVAMGAETRKSPITSSQFK-RSLELMKRTGACQPFVRCIKPBE 640
Qy      641 OKMPQPOAVVNLQRYSGMLETYRIRKAGYAVRPPQDFYKXKVMRLNA-----LPRD 697
Db      641 FKXPLPDLHLVCVROARTSGMMETIRIRAGYPIRSFVEGERKVLDPGKPAVYKQD 700
Qy      698 VNGKSTSLQLYDASNSEWQKTKVFLRESLQKLEKREBEVS-----742
Db      701 LAGTQORMAEAVLGHDDWQIGTKIKFLKDHMLLEVERDKAIDRVILLQKVRGPKD 760
Qy      743 -----HAMVIRAH-----VLGFL-----ARKQYKVLVYCVII 771
Db      761 RSNPLRLKSAATLQIRHMGHCRKQVYELIRGLFLQALHRSRKLKQYRLARORILEF 820
Qy      772 QCNVAPFLRRRFLUKKAIIVFOKQOLQIARVYROLAE--KREGEKKQBEBEKK 829
Db      821 QARCPAYLVKRAFRHMLAVITVOAVAKMIRRLRLRYEYORRLEBERMLABEELK 880
Qy      830 KRE-----EEREREREREAELRAQOEETRKQOELEALQSKQKAELETRELEKOK-- 881
Db      881 KREMAKKAKEBAEKHGERLAQAREDAE-----RELKEKEARKKELEQMEKARHE 935
Qy      882 ---ENKQVEEL-----RLEKEIELOKMEQOELSLTEASLOKLOERRDQ 924
Db      936 PINHSDMDVKMGFLGTSGSLPQOEGQAPSGFEDLERGR-----974
Qy      925 ELRLREBEACRAOERTLESINFEIDECRNTERLSGSEFSSSELAEACKEKNFNFS 984
Db      975 -REWEEDVDALPLP---DEDEEDSEYKFAFAATYFOGTTTHS-YTRRP--LK 1023
Qy      985 QP--YPEREVD-----GFEADDFAFKDSPNPEHGHSDQRTSGIRTSDDS 1028
Db      1024 QPLVHDDGDGDLAALAVITTLRFMG-----DLPERKX-----TAMSG 1064
Qy      1029 SEEDPYM-----NDTVVPTSPADST--VLLAPVQDSGSLN 1064
Db      1065 SEKIDVMTKIYETLGKTKYRELQALQEGEQTOLPEGQKTSVRHNLVHLTKKSKLITE 1124
Qy      1065 S-----SGSESTYCPQONAGDLPSPD-----GDVYQODDYEDGATIS 1102
Db      1125 EYTKRLNDESTVOQNSMLEDRTPSMLEKHLIIGNGILRLALREIYQISQK--LTH 1181
Qy      1103 GSVYTFNSYGSQMSPDYRCSYGTYNSS-----GAYRFSSEGAOSSPED 1146
Db      1182 NNS---KSYARGMILVSLC-VGCRAPSEKPYKILANFIHGPRPGYAPCEERLRTFPN 1237
Qy      1147 SEEDPSRPTDDELSYRDSVYSCVTLFPYFHSFLYMKGLMN-----SMKRRWC--- 1196
Db      1238 G-----TRTOPPSMLELQATKSKKPIMLP---VTFMDGTKTLLTDSATYARELCNALA 1288
Qy      1197 ---VLKD---ETPLMFRSKQBALKQGMH-----KKGGSSTLSRRMMKKW 1237
Db      1289 DKISLDRFGSLYLIALPKVSSLSGSDHVDALISQOYAKEOQA---OERAPMPL 1344
Qy      1238 FVLROSKLAMYFENDSEELKGVY-----EVRTAKB--IINDTTKENGIDY--- 1280
Db      1345 F-FRKEVTFPMNPSBDNATNLIYQOVVGVKFGEGYRCEKEDDLAELASQYFVDYGE 1403
Qy      1281 INADRTFLIAE-----SP-EDASQMFVLSQVH---ASTDQEIQEMHDEQANPON 1327
Db      1404 MLERLRLSLVPTIYIPREITPLKLEKMAQLAIANHKKGIYAOVRTDSQKVEDVNVYAR 1463
Qy      1328 AVGTLDVGLIDVCSADSPDRPNSFYIIRANRVLEGNADTPREMHMWTLLQSKQDTRV 1387
Db      1464 FWPMLPSRPFYEAYKFSGPLPKSDVIYAVNMTGVYFVEBOQV---LLELSPEEIMA 1518
Qy      1388 --EGOEFTY-----RGWLHKEVKNSPKMS--SLTKLRWAFVLTNSID 1426
Db      1519 VSSSRRCRVLLSLGCSDLCATCQOSRAGLITRAGPSCPCMSCRGTGMAAPSTLIATIKD 1578
Qy      1427 YKSSKRNALKLGTLVNSLCSVVPDEKIFKETGVM-----NTVYGRKHGCRYLTK-- 1479

```

```

Db      1579 EYFTSSNAEDIRDLVVTFL-----EGLRKRSKYVALQDNPNPAGESGFLSFAKGD 1631
Qy      1480 --LNEAT-----RMSVIONVDTKAPIDT-----PTOOLI----- 1509
Db      1632 LIIDHDGTGEQVMSGNANGINERTKQGRDPTCCVYVMPVTLPREIYALVMTTPDQR 1691
Qy      1510 QDI-----KENCINSNVQIYKRNPIILRYTHP 1538
Db      1692 QDVVRLQLRTAEBEVRAKPYTLEEFSDYDPRPPKHTLSVMVSKANGKRLMSHTREP 1751
Qy      1539 LHSPLPLPYGDINDLNLKDKGYTTLODEAKINSL-----QQLBSMDPIPIIQ 1589
Db      1752 LKQALK-----KIGSEBELSQEACMAVAVALKYNGDYPSKMRSVNE---LTD 1797
Qy      1590 GILQTHDLRLPRLBELYCOLIKQ--TNKVPRPGSVGNLYSQIITLCSCTFLPSRGIKY 1647
Db      1798 QIFEWALKAEPLKBEAVQIILKQLDNHIRISEERG---WELMLCTGLFPSPSILLPH 1853
Qy      1648 LKFLKRLREQFPGTEMEKVALFTYESLKK--TKCREFVRSRDEIALIHR-QEMTSTV 1703
Db      1854 VQRFLOS-RKQCP-----LAIDCLQRLQKALRNGSRKYPHLVVEBALQKHTQIPIKV 1906
Qy      1704 YCHGGGCKITTINGHTTAGEVVEKILRGLAMEDSRNMPAL-----PEY 1746
Db      1907 YFPDDTBAFEVBSSTKAKDCONIASRLILKSEBG-FSLPVKIADKVISVENDPFPDF 1965
Qy      1747 NGHYDKATESRTVADVLAKEPKLAATSEVGDLPMKFPF--KLACFLDTDNPV-KDSV-E 1802
Db      1966 VRHLTDWIKKARPIKDG-----VSLTYQVFPMKLM---TTVPQKDPAD 2010
Qy      1803 FAFMEQAEAVINHGHPAPEENLOVALA--RLQYLOGDYTLMAAIPLEBYVSLQRLK 1860
Db      2011 SIFHYQELPKYLGXKHTREBEVLQGLALYRVXF-BEDHSYPSPIKL----- 2059
Qy      1861 ARISQSTFTPCRLEKRTSIFLEGTIRASFRRGVSVRQVBEQOYMDMTKEVSSAR 1920
Db      2060 -----LR-----ELVPQDLIRQVSPDW-----K 2078
Qy      1921 ASIIDKMRFGQMNQOAMAKYMLIKEMPGVSTLFDV-ECKEGFPOELMIGVSADAV 1979
Db      2079 RSIYAVFNKAKGSKSEBKALFLLKILFMPFFGSAFBEVKQTTBENFELILAIANKGV 2138
Qy      1980 SVYKREGRPLEVFOYEHILSFGAPLANTYKIV-----VDBRELLFRTSEVVDVAKLMA 2034
Db      2139 SLIDRPTDILHTHPTISNWS--GNTYHITIGNLVRSKLLCETSLGKMDLLTS 2196
Qy      2035 YISMIVKKRYSTTSASQSSR 2057
Db      2197 YISQWL-----TAMSKQRNSR 2212

```

```

RESULT 6
A:Species: Homo sapiens (man)
C:Date: 19-May-2000 #sequence revision 19-May-2000 #text_change 02-Feb-2001
C:Accession: A59255; S71932; S60263
R:Chen, Z.Y.; Hasson, T.; Kelley, P.M.; Schwender, B.J.; Schwartz, M.F.; Ramakrishnan, M.
Genomics 36, 440-448, 1996
A>Title: Molecular cloning and domain structure of human myosin-VIIa, the gene product d
A:Reference number: A59255; MUID:97036886; PMID:8884267
A:Accession: A59255
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Accession: A59255
A:Reference number: S71932
A:Accession: S71932
A:Molecule type: mRNA
A:Residues: 96-334, 'T', 336-370, 'N', 372-388, 'VTG', 392-469, 'F', 471-563 <WEI>

```

A:Cross-references: EMBL:U1180; NID:G1235669
 A>Note: the sequence is revised in GenBank entry HSU39226, release 114, (PIDN:AA03679.1)
 R:Wells, D.; Blanchard, S.; Kaplan, J.; Gillford, P.; Gibson, F.; Welsh, J.; Mburu, P.; V
 Plet, D.; Munnich, A.; Steel, K.P.; Brown, S.D.M.; Petric, C.
 Nature 374, 60-61, 1995
 A:Title: Defective myosin VIIA gene responsible for Usher syndrome type 1B.
 A:Reference number: S60263; MUID:95174880; PMID:7870171
 A:Accession: S60263
 A:Molecule type: mRNA
 A:Residues: 96-157,199-255 <MEM>
 A:Cross-references: EMBL:U1180; NID:G1235669
 A>Note: the sequence is revised in GenBank entry HSU39226, release 114, (PIDN:AA03679.1)
 C:Comment: A alternate transcript from the same gene is shown in PIR:A59257.
 C:Genetics:
 A:Gene: MYO7a; USH1B
 A:Map position: 11q13.5
 C:Superfamily: myosin motor domain homology
 C:Keywords: actin binding; ATP; nucleotide binding; P-loop
 F/96-563/Domains: myosin motor domain homology (fragment) <MMOT>
 F/158-165/Region: nucleotide-binding motif A (P-loop)

Query Match 17.1%; Score 1833; DB 2; Length 2175;
 Best Local Similarity 26.2%; Pred. No. 5.8e-75;
 Matches 632; Conservative 368; Mismatches 802; Indels 614; Gaps 85;

QY 4 QGTRVWL-RENGQHPSTVNSCAEGIVFRTDYGVFTYKQSTITTH-----QKVTAMH 55
 DB 6 QGQVHMDLRSGGEFVPI-----GAVVTKLDCSQGVQVDEDEHMIQKNTAKPH 60
 QY 56 PINEEVDMASTLEHGSIMVNLFORAKNOIWTYIGSILASVNPYQPIAGLYPATM 115
 DB 61 PTVSHGEEDMIRIGDINEAGILRNLLIRYDHLIYTGSGILVAVPYQ-LISTVPEHI 119
 QY 116 EYSRHRLGLPHIFAIANEYCRLMKHRDNOCILIKGSGAGKSTETLILKFLSVS 175
 DB 120 RQTKKIGEMPHITAIANCTFNMKRNSRDCCIIISGSGAGKSTETLILQFLAALS 179
 QY 176 QQSLELSEKETSCEVERALLESPIIEAFGNATVNNNSRFGKQVQINICQKNIQGG 235
 DB 180 GQH-----SMIQVLEATPILEAFGNATINDNSRFGKIDIDHFNKRGALBEA 230
 QY 236 RIYDCLISQNRVVRONGERNYHIFVYALLAGLEHEEREFPYSTENYHYLNQSCVED 295
 DB 231 KLEQVYL-EKSRVCRALDERNHYVFCMLEGMSBQKKLGLGQASDVNYLGMGCITC 289
 QY 296 KTSIDESPREVITANDVWQFSKEEVRVSRLLAGILHGNIE-----FITGGAQVSRK 350
 DB 290 EGRVDSQEBANINRSAMKVLMTDTENWESKLAAILHGNLQYEARTENIDACEVLS 349
 QY 351 TALGRSAELLGLDPTQLTALTORSMFLGEEILTEPLANQQAADSRSLAMALYACCFEM 410
 DB 350 PSLATASLELVNPPMLSCITSRITITGETYSTLSRQALDVAAPKGIYGLFLFW 409
 QY 411 VIKKISR1--KNEDEPK---SIGLIDIFGFENFEVNHPEQFNINYANKEIQEYNNKII 464
 DB 410 IYDKINAAAYKPPSOVKNSRRSISGLDIFGFENFVANSFEQCLINFAHEHILQGFVRYH 469
 QY 465 FSLQLEYSRREGVWEDIMONGECLDI-EKKDLALLINEESHPPQATSTLLEKH 523
 DB 470 LKLEQSEYLDIESIDMLHIEFTDQDADMLANKPMIISLIDESRFPKGTDTMLHKN 529
 QY 524 SQHANHHFYVPRVAVNN---FGVGHVAGEVQYDVARGILEKRDTPFRDLNLALLESRF 579
 DB 530 SQHKMANANYIPK---NHETQGINHPAGIYVEYEGFLEKRDLDHDDIQLVHSSSN 586
 QY 580 DFIYDLFEHVSANNODTLKCGSKRR--PTVSSQFVDSLHSLMATLSSNPFVRCIK 637
 DB 587 KPIKQIF-----QADVAMGAETRKRSPTLSSQFK-RSLLEIMMTLAGACQPFVRCIK 637
 QY 638 PNNQKMPDQDAVVLNOLRYSGMLETVIRKAGVAVRRPFQDFYKRYKVLKRNLA--L 694
 DB 638 PNEFKKMPDLRHLQVROLRYSGMLETIRIRAGYPIRYSFVEFERYVLLPGVAPAYK 697

QY 695 PEDVKGCTSLQLDYDASNSSEWOLGKTIVFLRESLEOKLEKREEREESHAAVIRAHVYG 754
 DB 698 QGDLDNGTCORMAEAVLGGTHDDWQIDKTKIFLKDHHDMLBVERDAIIVDRVILLQKVRIG 757
 QY 755 FLARKQYRKVLVCVVIIO-----KNY----- 775
 DB 758 FDRSNFLTKKNAATLQHNMRGHCNCKYVGLMRGSLRLQALHRSRKLHQYRLARORI 817
 QY 776 -----RAFLIRRRFLKKAIVFOKQRLGQIARVYROLARK--REOEKKKQEE 826
 DB 818 IQFQARCRVYLVKAFRRLHMAVLTVQAVAKMIRRLHQRURAEYLRLEEKQALAE 877
 QY 827 EKKKEEEREREREREREALEAQOEEETRQOELALQSQKEALTRLEKQENKQV 886
 DB 878 EKLREKMSAK-----KAKEAEKQGERLAQLABEDAREL----- 913
 QY 887 EETLLEKIEIDLOKMQOELSTLEASIQKIQERDDELRLLEBACRAOEFLESUNF 946
 DB 914 -----KEKEAARAKE-----LLEQERARHBPVN-----HS 940
 QY 947 DEIDCVNRIEISLGGSEPSSELSAESACEEKPNFNFGQYPPEEVDGFEADDDAFKDS 1006
 DB 941 DNVDMKFGFL-CTSGGLPGQCGQAPSGFEDL-----ERGRREWEEDLDALAPL 988
 QY 1007 PNPSEHGHSDQRTSGIRTSDDSEEDPYMNDTVFPTSPADSTVLLAPSVDPSGLHNS 1066
 DB 989 PDEDEDESEYKFAFAFAAT-----YFGGT--TTHSYRRLKQP-----LLYHDD 1031
 QY 1067 GSESTYCNP-----QNAQDLPSPGDDYDYDD--YEGCAITSGSSVFS 1109
 DB 1032 EBDQALAAVMTILRFMGDLPPEP--KYHTAMSDSEKIPVMTKIYE--TLGKITYKR 1085
 QY 1110 NSYGSQWSPDYRCVGTNNSSGAYRFSSEGAQSPDESE-----DPDSRDTDELSY 1163
 DB 1086 ELQALQGEBAQLPBGQKSSVYRKLVLTLKKSRLTEVYTKLHDESTYQGSMLLED 1145
 QY 1164 R-----RDSVYSCVTLLPYHSEFLYMKGGLMNSMKRW----- 1195
 DB 1146 RYTNMLEKHLFIENGILRPLARDEIYQISQKLTN-----PEKSSVARGMILVSLCV 1199
 QY 1196 -CVLKDEFTFLMRSK-----QBALQGMHLKKGSGSTLRRMKKWPVLR 1241
 DB 1200 GGFASSEKFKVLYRNFHIGGPGVAPYCEERLRRTFV-----SGT--RTQPSMLBQ 1250
 QY 1242 QSK-----LMVFENSEEKLKGTVEVTAKEIID-----NTKENGIDITMA----- 1283
 DB 1251 ATKSKKPIMLPVTFMDDGTTKLTILDSATTAKELCNALADKISLKORFSGSLYIALFDKYS 1310
 QY 1284 -----DRTFHLIASESPDASQ-----W-----FSVLSQVASTDOEIQ----- 1316
 DB 1311 SLGSGSDHVMDAISQCEQYAKQGAQERNAPRLPFRKEVFTPMHSPEEDVATNLVIQO 1370
 QY 1317 -----EMHDEQANPONAVGTLIDVG-----LISVCASDSDPRNSVVIITA 1357
 DB 1371 VVRGVKFESEYRCEKEDDLAEIASQOYFVDYSGEMILERLNLV-PTYLPDR-----ITP 1424
 QY 1358 NRVLHCNADPTEPM-----HWITLLOR----- 1380
 DB 1425 LKTEKKAQALAAHKKGIYAQRRTDAQKVEDVSYARFKMPLFSFYENAYKSGSGL 1484
 QY 1381 SKGDT-----RVEGQEFIVRGMLHKEVKSPPKNSLKLKKRWFLTHNSLDYYS 1430
 DB 1485 PRNDVIVAVNMVTGVYFVDEQOVLLELSEFPEIMAVSSRGAKTAPSTLATIKDEYTF 1544
 QY 1431 SEKALKKIGTLVNSLCSVVPDEKIFKETGYW-----NVTYVGRKHCYRLYTK--LL 1481
 DB 1545 TSSNNEEDIRDLVVFPL-----EGLRGRSKYVALADPNPAGSESGFLSFAKGLLITL 1597
 QY 1482 NEAT-----RWSVIGNVTDTKA--PIDT-----PTQOLI-----ODI-- 1512
 DB 1598 DHDIOEQVMSGMAGINERTKQDGPFTDSYVYVPTTMBPRELIVATMTPDQRODV 1657
 QY 1513 -----XENCINSDVVEQIYKRNPIILRYTHHPLHSP 1542

Db 1658 RLLOQTAPBEVRAKPYLLEESYDYFRPPPKITTSRWVVSXARGDRLSMSTRBFLKA 1717
1543 ILPLPGDINLN-----LLKDKGYTLLOUEAIKFNISLOQLESMDPIPIIGIQ 1593
Db 1718 LAKLGLSELSEGOACLAIVALKVWG-----DYPSKRTSRVNELDQ-----IFGPIK 1767
Qy 1594 TGHDLRLPLDELVCQILQ--TNKVPHPSGVNLVSMQILTCLSCFLPSRGLKYLKPH 1651
Db 1768 A-----BPLKDEAVVQLKQTLTDNHIRYSEBG---WELLMLCTGLFPSPNILLPHVQNF 1819
Qy 1652 LKRIHQFPCTEKEKALTYESLK---TKCEFPSPDEIFALIR-QEWTSTYCYG 1707
Db 1820 LGS-RKQCP-----LAIDCLOQLQALNRSKYPHLVEVEALQHKTTQIFHKYFPD 1872
Qy 1708 GGSCKTINSHTTAGVEVKLIRGLAMEDSRMNFAL-----FEYNCHV 1750
Db 1873 DTDEAFEVSSTKAQFCQNIATRLILKSSG--FSLPVKIDKVLVSPENDEFFDVRL 1931
Qy 1751 DKAIERSVTAVDLAKFEKLAATSEVGDLPWKFPF--KLVCFLDTDNP-KDSV-EFAPM 1806
Db 1932 TWIKKARPIKQI-----VPSLTQVFPFKKLM---TTVPGKDPMAISIH 1976
Qy 1807 FEOAHAVIHGHHPAPEENLQVLAAL--RLQYIAGDYTLHAIPLPEEYVSLQRLKARIS 1864
Db 1977 YVQELPKYLRGYHKCTREEVQLGALIVYKF-EEDKSYFPSPKPL-----2021
Qy 1865 QSTKPTPBERLEKRTSFLBGLTRSPFTGSVVRQKVEEQMLDWIKEVSSANASTI 1924
Db 2022 -----LR-----ELVPDLIRQVSPDM-----KRSTIV 2044
Qy 1925 DKARKFQGNQEQAMAKYALIKEMPGYSTLFDVCEKSGFPOELMLGSADAVSVYKR 1984
Db 2045 AVFNKAGKSGKEAKLAFKLIFKWPFPGSAPFR-QTEPNPPEILLAINKGVSLIDP 2103
Qy 1985 GEGRPLEVPQYEHILSFGAPLANTYIV---VDERELLFETSEVVDAKLMKAYISMT 2039
Db 2104 KTKDILTTHTPFTISIMSS--GNTYFHTIGMLVRSKLLCETSLGYKMDLLTSYSQM 2161
Qy 2040 VKKRYSTRSASQGS 2055
Db 2162 L-----TMSKORGS 2171

RESULT 7
T25888
hypothetical protein T10H10.1 - *Caenorhabditis elegans*
C1Species: *Caenorhabditis elegans*
C1Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 08-Sep-2000
C1Accession: T25888
R1Nelson, J.; Langston, Y.
submitted to the EMBL Data Library, December 1996
A1Description: The sequence of C. elegans cosmid T10H10.
A1Reference number: Z20105
A1Accession: T25888
A1Status: preliminary; translated from GB/EMBL/DBJ
A1Molecule type: DNA
A1Residues: 1-2098 <NUL>
A1Cross-References: EMBL:U80848; PIDN:ABJ7988.1; GSPDB:GN00028; CESP:T10H10.1
A1Experimental source: strain Bristol N2; clone T10H10
C1Genetic1:
A1Gene: CESP:T10H10.1
A1Map position: X
A1Intons: 6/3; 43/3; 92/3; 154/2; 226/1; 280/3; 417/2; 472/3; 812/3; 853/3; 887/3; 978/3
C1Superfamily: myosin motor domain homology
F165-720/Domain: myosin motor domain homology <MMO>
Query Match 16.0%; Score 1716; DH 2; Length 2098;
Best Local Similarity 25.4%; Pred. No. 1.1e-69;
Matches 602; Conservative 364; Mismatches 806; Indels 596; Gaps 77;
Qy 4 OGTRVLRNGHFPSTVNSCAGIVFRTDYGV-----FTYKOSTITHQKVTANHPN 58

Db 6 KGDFTWI-EPK-----TEGSIPIGARVIDQHGRLKVIDLQNEQMLSADRRLMHTS 60
Qy 59 EGVVDMASTLELHGSIMYVFORYKRNQIWTYGISLTAISNPQIAGIYEPAITMBOY 118
Db 61 VQGVDMQOLGDFHESALIRNLFIYREKLIYATYGLIIVNPMIDIA-IYTADEIMY 119
Qy 119 SRRLGELPRHIFAIANECYCLMWRHNOGILIKGSEAGKTESTKLILFLSVISQGS 178
Db 120 KRKRIGELPRHIFAIADNAVTNMRREKNGQVVISGSGAGKTESTKLVLFPLATISGH 179
Qy 179 IELSLKETSCEYERILSSPIEMAFGAKTYVNNNSRFKKPVQVANIQKGNIOGGIV 238
Db 180 -----SWIEQOVLNANPVLNFAFGAKTIRDNSSRFQYIDVHNEGSGIEGAKIE 230
Qy 239 DCILSSONRVNRNGERNYHIFVYLLAGLEHEERBEYLSPTENYHILNOSGVEDXTI 298
Db 221 QYLL-EKSRIVYQSENENYHIFYCLLGLSREKSELELTADYTLIQKITYIAEGR 289
Qy 239 SPOESFREVTAMDVQPSKEVREVSRLAGILHGNIEF-----ITAGNAVSPKTA 353
Db 230 DDAADLAIIRSMRYLMIHQEIGSIFKLASLHIGNIRFRONTNDMMESVDVADPSTL 349
Qy 334 GRSAGELGLDPTQLTDALTQSMFLRGEIILTPLVQQAVIDSRDSLAMALYACSEWYIK 413
Db 350 VRIATLQLEQNLDIAITTKSLVTRERVISRLNGQAVADALAKAIVGKLFHIVR 409
Qy 414 KINSRI-KGNDFK-SIGILIDIFGENFEVNHFBQFNINANEKQIEXFNHGISPLE 471
Db 410 RVNDIIVRPSQRRSISILIDIFGENFESSFELCINFINETLQDFVHVHVFMEQKE 469
Qy 472 YSRGELVMDIDWIDNGECLDL-EKKGLALALINEESHFPQATDSTLEKLHSHOAHNH 530
Db 470 YDEHINNRHIFKVDNQATVDLIAQRPLNLSLIDEBEIPKRGDKTMLKLSHNGNE 529
Qy 531 FYVKEPVAVNN-FGYKHYAGVQVDVQGLLEKNDTPFDLILNLSRPFYIDLFEHV 589
Db 530 LYLQKSELOAFGTHAGVVFYNTFRLEKNDRSFADISLVLSISKMPFLALPDI 589
Qy 590 SSRNNQDILKSGSKRRPFTVSQPFVDSLSHLSMTLSSNPPFVNCIIPNMQMDQPDQ 649
Db 590 ----EYD-----SSRKKVTVGNQFR-RLSEQLMSQLTGHFFHICILPNEKRALVDR 640
Qy 650 AVLNQLRYSGMLETVIRKAGYAVRPFODFYKRYKLYAMNIALPE---DYRGKTSIL 706
Db 641 DLVLRQLATSGMETIKIRBSGYPIRHDYYPVFPYRVLVSSIQCPVARIIDLDHAKKIC 700
Qy 707 QLYDASNEWQIGTKVFLRSLEQKLEKREBEVSHAMVIRAHVIGFLARKQRYKLY 766
Db 701 HMLISTNADYQIGTKVFLKQKHDVLE-----QEVYRIKMD 737
Qy 767 CVVITQRYTRAPLRRFLHLKKAIVQKQLRGIARVYRQLAEKREBEKKQOEBE 826
Db 738 KAIVIQKVRRLVTRKDEKQKQAAVITQTMWRGDKQKRYQIIS-----783
Qy 827 EKKKKEBERERERERAEALPAQOEBETRQOELALOKSQAELTRLEKOKENKOV 886
Db 784 -----GSRLOAVLRSR-----LV 798
Qy 887 BEILRLKEIEDLQMKEQOELISTEASLQKQERRDQLRLLEBACRAAQEFLESJNF 946
Db 799 SHYQTLRTIITQF-----AVCRGSLVR---RQVEKRIKRGKAPLTVESSTASVIS 847
Qy 947 DEIDECVANIERSLSGSEFSELAESACEKPNPNFQPIYEEVVDGFEAFDDAFAFDS 1006
Db 848 DSHEELVHL-----FDF---LPSDGKDSGNB-----NDS 874
Qy 1007 PNPSEHG-HSDORTSGIRTSDDSEEDPYND-----1037
Db 875 ADSRRGYSRLHTSPVNPANIPRVDSYVDEBLSKYPGKKAATFPQAQATATYHVKPL 934
Qy 1038 ---TVVPTSPSADSTVLLA--PSVQDSGLNHSNGSBSTYCMPONAGLPPPDGDYDQ 1092

935 KVALHTHTPEPSAQLAALTAWTTLIFKNGDLADYKPG-----STNGSEVYDK 980
 QY 1093 -----DVEDEGATIS--GSSVTFSNSYSGQMSPPYRCVGTYNSSGAY 1133
 Db 981 TPVMIKLYATLGKKFSAHDLSEAMLSSEYGCATLTKKMKRKILSMTRKKGKINSDDIS 1040
 QY 1134 RPSSEGAQSSFEDESEDFDSRFTDDELST-----RDSVYS--CVTLPYHSPIL 1181
 Db 1041 SISDSYSSFFNAMLSE--NKPMTSLDPLHYIIGILREDBLREIYCOQLKQSLNPNPKL 1098
 QY 1182 YMGGLMNSMKRW-----CVLKDFTLMFRSKQKALQKQMKHKKGGSSSTLSR--- 1230
 Db 1099 -----SARCMWILSLCVCSPAPSERIFKTLF-----CFIRBGPAGTYGKYIE 1143
 QY 1231 -----RWKMKWFLVROSK-----LMYFENDESEKTL---KGTVEVTRAKEII 1269
 Db 1144 DRLRTOVNTGRORPSYVLOANKSOKPVYLAVTMDSGVKTLCDASATTAELCKQLA 1203
 QY 1270 DNTTKENGIDITIMADRTFHLIASPEDASQWFSVLQVHA--STDQETQE----- 1317
 Db 1204 EKVGLTNSFGFSLYIALFDKVSLSGCTDHVMDAISQCEQYAKEQGRQERNAPWRLFPK 1263
 QY 1318 -----WDEBAMP-----QNAV-----CTLDVGLI 1337
 Db 1264 EIFSFWHDEPRDDVSTNLIYQVIRGIKYGEYRCDEDELAALCAQOYYIDEGTMDVNL 1323
 QY 1338 DSVCASSPD----- 1347
 Db 1324 ENNLPSYLPDPFEMSGEMALEKQTOTIMQYRKKTGRBLSQIEVENVSVAKTWPIL 1383
 QY 1348 -----RPNSEVIITAN-----RVLHCNADTPREMHHTITLQRSKG 1383
 Db 1384 FSRFYEALKEAPPLPKNEVYIIANWTGVYVDDREHVMLEFSFPISTAYYKGRKSTT 1443
 QY 1384 DT-----RVSEGEPIVGMWLMKEVNSPKM--SSLKXKKRFFVLTNHSLDVYKSEKN--- 1434
 Db 1444 DTCTVTVVGDDEYTPQSPNADITNLIWFLBGLKGRSRYL---AISOQKDEKONPL 1499
 QY 1435 ALKLGTLVL-----NSLC--SVVPDEKIFKETGVWNTVYGRKHCVRLYT--KLHNE 1483
 Db 1500 EPEKGDLLIVNFENTGNTLTBESVKGENS-----RTCLFGLRAENVYVLPPLVLP 1551
 QY 1484 ATWSSVIONVDTKAPRIDTPTQOL-----IQDIKENCINSDV-----VBO 1524
 Db 1552 SKNTLOIFPKMDLSIDLFNNKQVTVVDYNAEPYLTLENFABEDNFNSQVAVSGQISLMT 1611
 QY 1525 IYKRNPL---RYTHHRLHSPILPLPYGDNLNMLND--KGYTTLDEDAIKTNSIQOE 1579
 Db 1612 LRKESQIECWRSREHIDQPLK-----KLNGREDACRG-----ALEIFAAL--MK 1656
 QY 1580 SMSDP-----IPITIGILQGTGHLRLPYADELYCOLIKOTNKVPHPGSVGNLYSMQIL 1631
 Db 1657 YMDDEBSKSRGLTHTLNDHFKLPI SMEALRDELQYQOLYQQLT--NPSIMSERGWEILL 1714
 QY 1632 TCIISCTPLPSRGLIKLYLKHRLKRIEOPFCTEMEKYAL--FTYESLKKTKCEFPVPSBD 1688
 Db 1715 WMATGFAPASALAKESHFLEK-----SRPHPIALDCQNMOKLAKGSGSKRYPHLY 1766
 QY 1689 EIEALILHR--QEMSTVYCHGSGCKTITNSHTAGVEVETLIGLMEBSRMNFALEPN 1747
 Db 1767 EYVAIHKHTQTOIHKVFPDNTDEALIVDSATAPRCHKIGYRLKGS--DGSFLP--- 1822
 QY 1748 GHVDAKIESRTVADVLAKEKLAATSEVODLPMKEVFKLYCFLDTDNPVKDSVEAFAMF 1807
 Db 1823 -----VKIKOKVLA-----VPES--EFFPY 1841
 QY 1808 EOAHEAVIHGHAPPE-----NLOVLAALRLQYQGVYTLHAALPRLAEVYSIQRLA 1861
 Db 1842 VRSLSDMVNTNATQDATWIPINYQVYFMRKLMY-----NFVAGADPOADIIPHNYQ--- 1894
 QY 1862 RISQSTKTFPCRLKRTKRTSFLSEGLTBSFRGSGVYVROKVEEQLMDWIKEE----- 1915
 Db 1895 ---ESQVLYGLYHAKTKTNDVIELAALILRSMTKDGKNAPLAQIPQLLDEITIPKDSLKMS 1951

QY 1916 VSSASASIIDKMKKEQGNQOQAMAKMALIKEMPGYSTLFDV--ECKEGFPQELMIGV 1974
 Db 1952 ASEWKTISNAVARIEHLKSDQAKIEPLNYICRMWTFPSGAFPPVQSYDNLNPDRLILAI 2011
 QY 1975 SADAVSVYKGGGRLEVPQYHILISFGAPLANTY-KIVY-----DERELLFTSEV 2026
 Db 2012 NOTGVNIYHLDYKNLLVQYPPFVIVCNMTS--GNTYFNMTVGNMLKGNKGKLLDITVGY 2069
 QY 2027 DVAKLMKAYISMIVKKRY---STRSAS 2051
 Db 2070 KMDDLTSTIISLISQNNHPSKTRKREVA 2097

RESULT 8

A59257
 myosin VIIa, short form - human
 C:Species: Homo sapiens (man)
 C:date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 01-Mar-2002
 C:Accession: A59257
 R:Chen, Z.Y.; Hasson, T.; Kelley, P.M.; Schwender, B.J.; Schwartz, M.F.; Ramakrishnan, M.
 Genomics 36, 440-448, 1996
 A:title: Molecular cloning and domain structure of human myosin-VIIa, the gene product d
 A:reference number: A59255; MUID:97038686; PMID:884267
 A:Accession: A59257
 A:status: preliminary; not compared with conceptual translation
 A:molecule type: mRNA
 A:Residues: 1-1203 <CHR>
 A:cross-references: GB:U55209; NID:g1613789; PIDN:AA050722.1; PID:g1613790
 A:experimental source: tissue type testis
 C:comment: A alternate transcript from the same gene is shown in PIR:A59255.
 C:genetics:
 A:gene: MYO7A
 A:Map position: 11q13.5
 C:superfamily: myosin heavy chain 95f; myosin motor domain homology
 F:68-729/Domain: myosin motor domain homology <MMD>

Query Match 15.9%; Score 1702.5; DB 2; Length 1203;
 Best Local Similarity 37.1%; Pred. No. 2.1e-69;
 Matches 382; Conservative 196; Mismatches 313; Indels 139; Gaps 24;

4 QGTRVWL-RENGQHPSTVNSGAEIVYPRDYGQVPTTKOSTTH-----QKVTAMH 55
 Db 6 QGDHYWMLRSGQEDVPI-----GAVVYKLCDSGVQVVDDEDNEMWISPONATRIKPMH 60
 QY 56 PNEGVDMAELTLEHGSITVNLFORYKRNQVITYGSLASVNPQPLAGLEPATM 115
 Db 61 PTVSHGEDMTRLGDLENAGILRNLLIRYRDLITTYGSLIVAVNPQ-LLISTSPEHI 119
 QY 116 EGYSRRLGELPPHIFAIANBECYRCLMRHNOCTILNGESGAGKTESKTLKEFLSVIS 175
 Db 120 RQYTNKIKGEMPHIFALADNCYFMKRNNSDQCIISGESGAGKTESKTLILQFLAIS 179
 QY 176 QQSLESLKEXTSCYERAILBESSPIMEAFGNKATYVNNNSRFGKFFVOLNIOCKGNIOGG 235
 Db 180 GQH-----SWIEQOVLEATPILEAFGNKATIRNNSRFGKYIIDIHFNRKALIGA 230
 QY 226 RIVDCILSNRVNQNGERNYHIFVALLAGLHEHEEPEFLSPREMYHYLVNQSGVED 295
 Db 231 KIEQYLL-EKSKVCKQALDERNYHVFYCLBMSMSDQKKKGLGASQSYNVLAMNCITC 289
 QY 236 KTISQESFREYITAMVQFSKEEVRVSRLLAGILHIGNIE-----FITAGQAQVSPK 350
 Db 290 BERVDVSGEYANIRSMKVMFLFDTEINWISIKLAAIILHGLNQLYEARFFENLDACEVLFS 349
 QY 351 TALGRSABLGLDPLQTLDTALTONSMPLRGEIILTPLVANQOAVRSDSLAALVACCEW 410
 Db 350 PSLATPAALILEVNPDLMSCLTSRLITRGEVTSPLSREBALDVYDAFVVGIGYGRLEPVM 409
 QY 411 VIKIKINSRI--KGNDFK-----SIGILDIPGENEVNHPPEFNTVYNEKLOEYFNKHI 464
 Db 410 IVDKINAAIYKPPSDVKNRSRSIGLDLIPGENFAVNSFQOLCLINFAHEHLOQFFVZHV 469

QY 465 FPLEQLEYSREGIWMEDIMINGECLDI-EKKGLALINEBSHFPOATSTLLEKIH 523
 DB 470 LTLDEEYDLESIDMIHIEFTDNOALMIANIPMIISLIDESKFPKGTDTMLHKIN 529
 QY 524 SOHANNHFPVKRVAANN-----FGVGVAGEVQYDVRGILEKRDPRDMLNLRBSRF 579
 DB 530 SOHKLNANTIPPK-----NNHETQGINHFAGIYIYETQGFLEKRDLDHDDILQLVHSSN 586
 QY 580 DEIYDLEFHVSSRNNDTLKCSKHRR--PTVSSQFVDSLHSLMATTSSNPFVRCIK 637
 DB 587 KTKIOLF-----QADVAMGAEPTKRSPTLSSQPK-RSLBLMKTLLGACQPFVRCIK 637
 QY 638 PMWQMPQDPODAVLINQLRYSGMLETVRIRKAGAVRRPPODFYKRYVLMENLA--L 694
 DB 638 PNEFPKPYMLFDRHLCYRQLRYSGMETTIRIRAGYPIRYSFVEFERYVLLPGVPAVK 697
 QY 695 PBDVKGCTSLQOLVYASNSWOLGKTVPLEBSLEQLEKREBEVSMAAVIRAHVIG 754
 DB 698 QGDLRKTCORMABAVLGTDDMOIGTKTFLKDHMDLLEVERDKAITDRLVILQVYIKG 757
 QY 755 FLARKQYKVLVCVVIQ-----KNY----- 775
 DB 756 FDRNSFLKLNAAATLIQHMNGHCKRYGIMRLGSLALQALHRSKRLHQYRLARQRI 817
 QY 776 -----RAFLRRRFLHKKAAIYFQOKLRQOARVYRQLLAEK--REQEEKKKQEE 826
 DB 818 IOFARACRAYLVKAKFRHLMAVLTVQAYARGMIAARLHQRLAEYMLRLAEKQOLABE 877
 QY 827 EKKKE-----EEERERERREALRKOEEETRKQOELALQSKQEAULTRELEQ 880
 DB 878 EKLREKMSAKKAKEERKRGQERLQLABEDAE---RELKEKKAARKKELLEQMEYA 932
 QY 881 K-----ENKQVEIL-----RLEKEIEDLQRMKQOELSTEASLQKQER 921
 DB 933 RHEPVNHSMDVMKMPFLGTSGGLPGQESQAPSGFEDLRGRRE---MVEDDLAALPL 988
 QY 922 RQDELRLLEE 931
 DB 989 PDEDEDLSE 998
 RESULT 9
 AS9233
 myosin VII-like protein - fruit fly (Drosophila melanogaster)
 C:Species: Drosophila melanogaster
 C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 17-Nov-2000
 C:Accession: AS9233
 R:Anderson, J.B.; Yamashita, R.A.; Sellers, J.R.
 submitted to Genbank, February 2000
 A:Description: Complete cDNA for an unconventional Myosin (Class VII) in Drosophila mel
 A:Reference number: AS9233
 A:Accession: AS9233
 A:Status: preliminary; not compared with conceptual translation
 A:molecule type: mRNA
 A:Residues: 1-2121 <AND>
 A:Cross-references: GB:AF233269; PIDN:AAF34810.1
 C:Genes: Myo28B1
 A:Gene: Myo28B1
 A:Cross-references: FlyBase:FBgn0040299
 A:Map position: 2, 28B1
 C:Superfamily: myosin motor domain homology
 F:69-725/Domain: myosin motor domain homology <MMO>

Query Match 15.8%; Score 1697; DB 2; Length 2121;
 Best Local Similarity 25.9%; Pred. No. 8.1e-69;
 Matches 588; Conservative 362; Mismatches 847; Indels 484; Gaps 83;

QY 5 GTRVLMRENGOHFPSTVNSCAE-----GIVFRITDYGQVFT-----YKQSTITHQKY-TAM 54
 DB 9 GEYVWVWPQ-----NTTSEPAVPFGARIVRTETQTLLVCDNRKQFQWVAGDVLKAM 60
 QY 55 HPTNBEQVDMAASLTLEHGSIMYNLPRYKKNQIVTYIGSILASVNPQPIAGLYEPAT 114

DB 61 HITSQEDVEDMITLGLQEYTLRLNQLNRVAKOLITYTYSMLVAINDYQ-ILPIYTNRE 119
 QY 115 MEQYRRHIGELPHIPIAINECYRLKWRDNOICILIKSGSAGKTESTKILKFLSVI 174
 DB 120 IQLYKNSLAEPLPHIPIAISDNAQRLORLKENQCVVSGSGAGKSTSTLILQYLAII 179
 QY 175 SQQSLELSLKERTSCVERAIILESSPIMEAFGNAKTVYNNNSSRFQFVQVNLICQKNIQ 234
 DB 180 S-----QKHSIEWEQIIEANPIMEAFGNAKTVRDNSSRFQYIEIRPTGALQ 230
 QY 235 GRIVDCILSSQNRVYRQNGERNYHIFALLAGLEHERREPYL--STPENYHYNQSGC 292
 DB 231 ARIQOYLL-EKSRIFYQSDERDNYHIFYCMLAGISTAREBKQOESPSQYHILAQGC 289
 QY 293 VEDKTIQSQSFREYITMDVQFSKEEYREVSRLLAGILKGNIEFITA-----GGAQV 347
 DB 290 FTLPRGAKAPADIRAAKMLSPFEPEVMSILSLAAILHGNLRFATAVANLATMEI 349
 QY 348 SFTALGRSABILGLDTQLTALTORSMPLRGEIITPLVQVQAVDSRSLAMALVACC 407
 DB 350 DDTPLQGVAAQLGIPISALNAALTQRTIFVHGEHVTTLSKKAIEGRDAFVKSLYDGI 409
 QY 408 FEWYIKKINSRLKMGED--FKSIGILDIFFGENPENVHFEQPNITVYANEKLYOEYKNGIIF 465
 DB 410 FVRIYRINETINKQVDQPMNSIGVLDIFGENPDNNSFEDLCINYAENMLQQPFVGHIF 469
 QY 466 SLEQLEYREGIWMEDIMINGECLDI-EKKGLALINEBSHFPOATSTLLEKHS 524
 DB 470 KMEQDEYQNEHINMOHIFQDNQOQLDLIGKPMKMSLIEBSKFPKGTDTLLEKHLV 529
 QY 525 QHANNHFPVKRVAANN-----FGVGVAGEVQYDVRGILEKRDPRDMLNLRBSRFPIY 583
 DB 530 QHGNRSIYKGTQTSFGIRHAYAGVYVMPDLGLEKGRDPSFGDLATVLQYSTNKYLV 589
 QY 584 DLFEVSSRNNDTLKCSKHRRPPTVSQFVDSLSHSLMATTSSNPFVRCIKRMQYM 643
 DB 590 DIFPH---EMPMDTK-----KQPTLCVKFR-NSLDMRLRTLSQHPYFIRCIKENBYKE 640
 QY 644 PDQPOAVVNLQRLSGMLETVRIRKAGVARRPPODFYKRYVLMRYMALPE--DVGK 701
 DB 641 PGNFDELQVQLRSGMLETARIRAGYPIRHAIVARVERLRLVPPVGLQCCDKKL 700
 QY 702 CTSLLQLYDASNSEQOLGKTVPLEBSLEQLEKREBEVSMAAVIRAHVYGLFARKQY 761
 DB 701 ARQICEVALPADSDQYQKTLFLRDEBDABELEQSLMLKSYITQIRGIRVLPFRYVM 760
 QY 762 RKVLVCVITIQNTAPFLRRRFLHLK-----AAIVFQOI----- 798
 DB 761 KYREALITVQRYMGRLOERKRYQVMRQGFRLGACIAOQLTTFYVRCRTIKQALS 820
 QY 799 RGQIARRVYRQLLAEKREBEKKQEEKKREBEER-----ERERERREALRKOEE 854
 DB 821 RGYLVKDFQKTLERRKQNDLKEELIKLAKMEABEELRLQLEKQERORQOQKR 880
 QY 855 TRQOELBALQSKQEAELTRELEKQENKQVEEILRLKEEIBDQRMKQOELSTEAS 914
 DB 881 LQEQRLKA-EAARNALAMAQKRTKPYKQ-----EAPKAPYQIANSRLPPPT 933
 QY 915 L---OKLOERDQELRL-----EACRAQPFLEJLNFDEIDECVNIERSL 960
 DB 934 LIVAAPLRTPASAVTRINTIPESFGTIDVSSKQWVDVFRFLN-DEPDALRLKLNIS 992
 QY 961 SGGSEFSELBASCEKPFNFQO-PYREBEVDEGFADD--AFKQSPNPSHGSHDQ 1017
 DB 993 SGD---TRLPKSPNNIDTSDFLKYAATYFGGATQHEKRPKLKSLKHEHPIDEM 1049
 QY 1018 RTSGL-----RTSDSSEEDPYMNDV-VPTSPASDSTVLLAPVQDGSGLHNS-SGES 1070
 DB 1050 ASKALWILLTLFMDG-----LPDVYSSPTLHVPFNDNLMS-----DLASLNTSDSYKP 1098
 QY 1071 TYCMFQNAQDLPSPGDYDYQDDYEDGATITSGSVTFPSNSYGSQWSPDYKCVCTYNS 1130
 DB 1099 RLTVQOSORRIPKP-----LASGEKE--AQEFYQHM-----LNVFTSHLE 1136

```
QY 1131 GAYFSSSEG-AOSSPEDS--EEDFDSRFDTDDELXYRSDSVSCVTLPYF-----HS 1179
Db 1137 KIHFIIGHGIIKNSLRBEILAQCKOYLNPSSRSYRGWLLSLCISCPREKEREPLH 1196
QY 1180 FLYMKGG-----LMSWKCR 1194
Db 1197 RSFMRKGGTQALQATPSIQRLERTLVNGPRCPSPLFELNARGRNPLKLDIHLMDCQHR 1256
QY 1195 WCYLDKETFLMFRSKOEBALQO-----GWLHKKGGGS-----ST 1227
Db 1257 LQYDAAST-----AREAVQLOQSMGLTDTFGFGLVMSLNGKLMPLGAGQEVDAISE 1310
QY 1228 LSRNWKRRV--FVLRSKLMYEENDSEK-----LK-GTVEVTKAEI1DNT 1272
Db 1311 CEQRQLDAPKLYIRKEMFATWYDPSMDPKATQOLYKQILNGKCEGYSRSEKDIAMVC 1370
QY 1273 T-----KENGIDIMADRTFHILIESPED-----ASQFVLSQVHASTDDEIDQMD 1320
Db 1371 ALACFVEYGGELIRLKPSEITFVPSDLAPGERAIEMNSRLIAATYKESVYKEEOND 1430
QY 1321 EQANPONAAGTLDVLGLDSVCASDSPDRP-----NSFYIIPANRVLHCNADTPEEMHWIT 1376
Db 1431 LLEAQKRAK-----EDICLFAHLSWPRHSHRLBEV-----RKEGPKLQSDML 1475
QY 1377 LLORSKGDTRVEQEERIVRGMLHKE-----VKNSPKMSLKLKRWFFVLTHNS----- 1424
Db 1476 LGINSAGLFLIDETEOVLASCSEFVYLKHYVESDDKLVHVTFOHNVFVLQCSGSAQDANEV 1535
QY 1425 ----LVIYSSSEKALKLTIVLNSL--GSVPPDEKIFETGYVNVVY--GRKCYL 1476
Db 1536 INYMLNLRORSSYGAVALPVPVGGDLEDCLVLPBGDIIEEAGVTAQALMAGNAQDCYR- 1594
QY 1477 YTKLNEATRWSSVIGNVDTTKARIPRTOQLIODIKENCLN----- 1518
Db 1595 --GCVA--GOMGFLGNVRVLTLYKPSSEKQDIIIEGAFQEPKPYTPRANYSRRQHN 1650
QY 1519 -SDVNEQIYK-----RNPLRYTHNPLNSPLLPBYGIDINLNLKDKGYTLQDEAIKI 1571
Db 1651 ISQLASHFRPRLDSDKAPLSKFSPEPLKAPLL-----KAVVAV 1689
QY 1572 FNSLQO-----LSMSD-----PIP---IIGLIGTGHDLRPLRDELVQOLIKQTN 1614
Db 1690 PPLFQALVNMHHILKRYMGDIARSNLPVNTDLIFQPALQ--HPL--LCDELYQOLMKQS 1745
QY 1615 KYHPSPVGNLYSMOILTCSTFLPS-----RGILKYLKPH-----LKRIREQPGH 1662
Db 1746 --DNPSSESEKRMDDLILATGLVAPSVLVMEELIILMRADALADACLKRLKSLAOG 1803
QY 1663 EMEKYA-LFTYESLKTKTC-----REFVPSRDEIALIHRQEMTSTVYCHGGSCKITTN 1716
Db 1804 ORKKAHLIEVEGIIQ-RCILHYKITYFPD-DTVEA-----FEIE 1841
QY 1717 SHTTAEVVEKILRGLAMEDSRMFPALFEYNGHVDAIESRTYVAVLAK-----FEKLA 1772
Db 1842 SHTRGAEILADIAQRIELK-SPVGYSIPLKTGDRVYVAMEEVEFVFPIQLIYMLRQORT 1900
QY 1773 TSEVGLPMKFYFKLVCPLDTPVPSKV--BPAFMEOAHNAVINGHHPAPENLQVA 1830
Db 1901 IRSISGQVQJHMKRLML--NNHPSGDLNGDMIFSYPOELAKYLVGGYPIDECOASRLA 1958
QY 1831 ALRLQYLOGDYTLHAAIPRLAEVYSIQRLKARISOSTKTFTPCERLEKRTSFLBGTLLR 1890
Db 1959 IL-----VYSADHDVSLQRLP----- 1974
QY 1891 SFTGSGVROKVEEOMLDMWIKESVARSASIIIDMKRFQGNQGNQAMAKWALIKEMP 1950
Db 1975 ----EVLRLRLPE---DLIPLOQVAERQOILPVNR-DHLEDAHALIFLOEISHTA 2024
QY 1951 GYGSTLEFDV-ECKEGGFPOELMGVADAVSVYKRGEGRPLVFOVEHLISFGAPLANTY 2009
Db 2025 CRGOSTFFVYKQONDALPETILAINSTGFRHMDPTTKELINSYEISQ-LGIWSSGKNHF 2083
```

```
QY 2010 KI-----VDERELLEETSEVVDVAKMKRAYI 2036
Db 2084 HIRFGNIGASKYLCTSTGYKMDLLASYV 2114

RESULT 10
A59295
unconventional myosin-15 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 08-Sep-2000
C:Accession: A59295
R:Liang, Y.; Wang, A.; Belyantseva, I.A.; Anderson, D.W.; Probst, F.J.; Barber, T.D.; Ma
an, T.B.; Frideil, R.A.
Genomics 61, 243-258, 1999
A:Title: Characterization of the human and mouse unconventional myosin XV genes responsi
A:Reference number: A59295; MUID:20021762; PMID:10552926
A:Accession: A59295
A:status: preliminary; not compared with conceptual translation
A:molecule type: mRNA
A:Residues: 1-3511 <L1A>
A:Cross-references: GB:AF14095; NID:96224684; PIDN:AAF05904.1; PID:96224685
C:Genetics:
A:Gene: MGI:Myo15
A:Cross-references: MGI:1261811
A:Map position: 11:33..9
A:Superfamily: myosin motor domain homology
F:1209-1871/Domain: myosin motor domain homology <MMO>

Query Match 14.8%; Score 1593; DB 2; Length 3511;
Best Local Similarity 24.8%; Pred. No. 8,3e-64;
Matches 609; Conservative 373; Mismatches 836; Indels 642; Gaps 81;

QY 59 BEGVDDMASITELHGGSIMYVLFQYKRNQITWYIGSLASVNPYQPIAGLYEPATMEQY 118
Db 1205 EDGVDDMQLEDQLOTTVLANLKTFRNLITYTIGSLIVSNPRMRA-ITGPEVOQCY 1263
QY 119 SRRLGELPPIHFAIANECYRCLMRHNOCLIKGESGAKTESTKILKFLSVISQOS 178
Db 1264 SGRLGENDPPIHFAIANLAFAMLDKONQCVIIGESGSGKTEATKILRLAMNQR- 1322
QY 179 LELSLKETSCYERAILTSSPIMEAFGNKATVYNNSSRFQVQNLNCOGNIGGRIV 238
Db 1323 -----BDVMOQILBATPLLEAFGNKATVYNNSSRFQVQNLNCOGNIGGRIV 1373
QY 239 DCILSSONRVVRONGERNYHIFYALLAGLEHEEREFEYLSTPENYHYLNOSGCVEDKTI 298
Db 1374 SOYLEKSRIVQAKENNNYHIFYELLAGLPRLQRLQASLSQEAETVYVYLNOSGNEINGK 1433
QY 299 SDQSFREYITAMVQPSKEBEVRSRLAGILHLNIEF-----ITAGGAQVSKTAL 353
Db 1434 SDADFFRRLAMEVLGFTSEDDDSIFRILASILHLGNVYFEKHEETDAQEVLASVARSBI 1493
QY 354 GRSAEILGLDPTQLDALTORSMPLRGEIILPLVNOQAVDSRSLAMALVACPEWYIK 413
Db 1494 QAVALLQVSPGLQKALTFTKTEIRKIFPLPLVESAVDARDIAVLALFLGMLIT 1553
QY 414 KINSRIKNEDEPKSIGILIDIFGFENFEVNHFEQFINVANEKLOEYFNKHFISLEOLEYS 473
Db 1554 RNALVSRKQDTLSAIIIDYGFEDLSFNSPQOLCINANENLOVLFKLYIPOEQEBEYI 1613
QY 474 REGLVEMEDIDWINGECLDIE-KKGLALLINEBSHPQATSDTSLTEKLSQHANHFY 532
Db 1614 REQMDREIAFADNPQICINLISLKPYGILRIIDDCPCFPQATDHTFFLOKCHYHGANPLY 1673
QY 533 VKPRVAVNPFQVKYAGVOYDVKGILEKNNDTPRDDLLNLRBSRPFITDLEFHVSR 592
Db 1674 SKPKMPLPEFTIKHAYGVTYQVNHKEFLDKNDQVQVDLDFVHSRTGVVAHLFSSHAQ 1733
QY 593 NNQDTLKCGSK-----HRRPTVSSQPKVDSLHSLMATLSSNPPFRCIKPMNQKPPQFD 648
Db 1734 TAPPLLGSSSITRLRYKATYAAKQ-OSLDDLVKMERCHNPLFPRCLKPRHKEPGLFE 1792
QY 649 QAVVLNQRLYSGMLETVAIRKAGYAVRRPFDQFYKRYVLMR-NIALPBDVYRGKCTSL- 706
```

Db	1793	PDVMAQDARSVGLVETVIRIRKEGFRVRLPFQVIFIDRRCUVALKLTANPAD--GDMCVSLLS	1851
Qy	707	QIYASNSEWOLGKTQYFPLRESLEQKLEKREEVSHAAVITRAHVGLFLARKOYKULY	766
Db	1852	RLCTVTPDMYVAVGSIKFLKEHLQOLLESMBERVOQNAAALTLOREYLGFFLOHFHSILRR	1911
Qy	767	CVVILIOKNYRAFLLRRRFLHKKAAYVQOQOLRQOILARRVYROLIAEKBQ--EKKKQOE	825
Db	1912	KIILQSBARCFLLAORIQOMRQSLKFRSLVHTVYNNRRRLKLRABORRRRQOAMLRBQ	1971
Qy	826	EKKKKREEREREREREREREAELRQOEERTKOQOLEALQKQOEALY--RELEKQKE	882
Db	1972	EELSKREYV-----PVNHELVPA-----EVAGLQAAAGKLSGSEPRVAUVRAP	2015
Qy	883	NKQVEILRLKEJEDLORMKEQOELSTEASLKLQERRDOELRLE--EESCRAOEPL	941
Db	2016	RLQABPCVTLRLDINNYP--MAKFIKCFEKESFCMLTVPLMPLTRLRVEHNAEAIYVFK	2074
Qy	942	ESLNPDEIDECVNRNIERSLGSSEFSBLASAEK-----PR--FVFSQPY--	987
Db	2075	LILRFPMG--DPLHGTQEMIIGNIYVHOGLVPALRDELIAQLANQWNRPNAYVSKRGWL	2133
Qy	988	-----PEEYDE-----GEADDDAFKDSPNSEKSHSQPQSGIRTS	1025
Db	2134	LLAACLGFAFSPHLDKFLKFVSDYGNQGA-----VQNRILQAMSGAART	2183
Qy	1026	DDSEEDPEYMDVTPVTPSPADSVLLAPVQDSSSLHNSGSGSTVCMQNAQDLPSPD	1085
Db	2184	PPPIQO-----LEMTAIAQEKVAMALDVSCFNQDQFQSPVHTWSTGEAV-----AGDILKHR	2233
Qy	1086	GDYVDDDDYDDGALTSSSVTFNSYSQMSPP-----DYKSV-----GTYN	1128
Db	2234	G-----LADG--WRGMTVAMKN--GVQMAELAGHDYVLVDLSDELRLDPFROKSYF	2281
Qy	1129	SSGAYRFSEGAOSFEDESDPQSPFPTDDELSTRDSVYSCVTLPRFHSFLYMK--GG	1186
Db	2282	IVGA-----EGPLAGRGDTRGVFGNCWDSDEDTPLRPOPQHVAKMPLDQYCSHKEDGT	2336
Qy	1187	LMNSWKRRM-----CULKQETFLMFPSKO	1210
Db	2337	NGERTAQGWTSNRQAVDSIGESTVPRPRLDGYLDSLPVVLACQADALEKTAIAYMK--	2395
Qy	1211	EALKGWLLHKHGGGSSST--LSRRNWK-----KRVFLYROSKLMTYEND	1251
Db	2396	-----GGGQPGGGGGSTEDTSRRPPEKCLKPIQGLDASTIALQOAFIRQAVLLAREMT	2450
Qy	1252	-----SEBKIKGT-----VEVR-----TAKETIIDNTTKEN-----	1276
Db	2451	LQALLAQOQPSAITSRRPOLPERPLAPERAPRTVVGTEPPAPVVLVRPPQSWADGSYAKA	2510
Qy	1277	-----GIDIIMADRTF--HLIAESPDAQMFVLSQVH--ASTDOEIQEM-----	1318
Db	2511	PKIESKPVAVETILADQWTAPESIASPELVN--YSTLNSHFPPQTOIRBITIQOYKOPR	2568
Qy	1319	-----HDE-----OANQNAVGTLDVGLDSV--	1340
Db	2569	WAGHPPEARTRDGGKVFRRPRPBHEALMILKQKTOQLAVBGTQVSREAVAMVPRVTSAP	2628
Qy	1341	-----CASDSPDRNSYV-----ITAN-----	1358
Db	2629	RPCMGPTVPOQSRSLERPEDEPVQOLRLVAPNFYGOIDIPWRIFLRKEVYPRDNTSHP	2688
Qy	1359	-----RVH-----CNADTPEMHMTLLLORSKGPTRVGEQEFIVGMLHKEVKN	1404
Db	2689	VQDLDFQOILHDTFSEACLAISBERLQMKALFRAGQNDIQO-----RPLVTSYSKR	2740
Qy	1405	S-----PKWSSL-----KLKKWFVL--	1420
Db	2741	AALSWARDSWEIYFRLFPFAMGVSQTVQOILAVSHTGKLLQMWGSKSARRLRVLICAY	2800
Qy	1421	-----THNSLDYKSSSEK-----NAKIKGTIV-----LMSLGSV	1450

Query March 14.8%; Score 1589.5; DB 2; Length 3530;
Best local similarity 25.1%; Pred. No. 1.2e-63;
Matches 623; Conservative 346; Mismatches 818; Indels 697; Gaps 88;

QY 59 EECVDDMASLTJLHSGSIMYNLFORYKRNQIMTYIGSIILASVNPYOPINAGLYEPATMEQY 118
 Db 1221 EDCVEDMTQLEDOETVLSNLKIFREBNLIYTYIGSILVSNPYQ-MFGIYGEQVOQY 1279
 QY 119 SRHJLDELPHIATANECEYRCMLKRDHNOCLIKESGAGKTESKTLIKELSVISQOS 178
 Db 1280 NGRALDENPHILAVANLAFKKMLDAKQOCIIISGSGGKTEATKLILRYLAANNOK- 1338
 QY 179 LELJLKEKTSCEERAILSESPIMEAFGNACTVYNNNSRFGKFOVONICOKNIGOGRAVY 238
 Db 1339 -----REWMOIKR-ILFATPLLESFGNACTVYNNNSRFGKVEIFL-EGGVISGALT 1389
 QY 239 DCILSSQNRVVRONPBERNYHIFYALLAGLEHBEREPYLSTPENHYILNOSGCVEDKXT 298
 Db 1390 SYLLERKSRIVFOAKNERNYHIFYELLAGLPAQLROAFSLQEAETYYILNOSGNCISIAQ 1449
 QY 299 SDOESFREVTIANDWQFSEKEEYREVSRLLAGLHGNIEF-----ITAGAOVSFKTL 353
 Db 1450 SDADDFRRLLAAHEVIGFSSSEDDOSIFRLASILHLGNVFEKXETDAOEVSASVABEI 1509
 QY 354 GRSAEILGLDPTOLTALTORSMFLRGEELTPLANVOAVDSRSLAMLYACCFEMVIX 413
 Db 1510 QAAAEILLOISPEGLQKAITFKYETREKIFTPLYESAVDADALAKULYALLFESWILT 1569
 QY 414 KINSRIKGNEDFKSIGLIDIFGEFENFEVNHFEQFNINYANEKLOEYFNKHTFSLQLEBYS 473
 Db 1570 RVALVSPRODTLSIILDIYGFEDLSFNSFEOLCINYANENLOYLFNKIVFQEEQOEYI 1629
 QY 474 REGUVEDIDMINGECDLIE-KUGLALINEBHFPOATSTILEKLSHANNHY 532
 Db 1630 REGIDMOETIFANQPRINILSLKPYGILRIDDOCCFPQATHTLQCKHYHGNPNLY 1689
 QY 533 VKPRVANNFEGVAGEVOYDVRGLEKRDTPRDDNLILRESRFDYDLF-BHVS 591
 Db 1690 SKRKMPLPETTIGYAGKTYOVHKLFDKXHDQVRDVIDLFRKSTRVVAHLFSSHAQ 1749
 QY 592 RNNQDTLKCSG---KRRPTVSSQFVYDLSHLSMATLSSNPFVACIKRNQKMDOPD 648
 Db 1750 AAFORLGKSSVTRLYKATVAAKFO-QSLDLVEKMERCNPLFMCLKNHKEGFLPE 1808
 QY 649 QAVVNLQALYSGMLETVRIKAGYAVRRPRDQYKTYKULMRLALPDEVGRK---CTSL 705
 Db 1809 PDVVAQLRYSGLVETVIRKEGFPVRLPFOGFIIDYCCV---VALKHLRANGDMCVSV 1865
 QY 706 L-OLYASNSEWOLGKTYFLRESLEOKLEKRESEVSHAAMVIRAHVIGFLARKOYK 764
 Db 1866 LSRLLCKVMRYVGVUSKFLKHLVQLESKREHYNLALTLQCLAGFTIKRFRSL 1925
 QY 765 LYCVVILQKNYRAFLRRFLMLKKAIVFOKQLOIARVYRQDLAEKREQ-EKKKQ 823
 Db 1926 RHHIILLOSARGYLARQYQWRSLVFRSLVHAIVSRRLYKLRARERCQVEGALLM 1985
 QY 824 EEEKKRR-----EEBREERERRRAELRAQOEEST 855
 Db 1986 EOEELSKREVAVAGHLEVPALAGLLOAVAGLGAOVPAVPTBRLOAEPVTLPLDI 2045
 QY 856 RKQOELEALQOSKE---AELTRELE-----KOKENKQVEETILREKEIEDIORMKEQ 906
 Db 2046 NNYPMAKFVQCHKEKAPFGLTYRPLRPLQOLRAHHAARVASFUKILAFMGDPRLHNGR 2105
 QY 907 ELSLSTASLOK---LOERDQELRLEBEACRAAOEFLFSLNDEIDECVRNIERS--- 959
 Db 2106 ENIFGVYIVQKGLAVELDEILAOQLANQVH-----NNAHAHAERGMILL 2151
 QY 960 ----LSGGSFESSLASACEEKNFNSQRYPREEVEDGEFADDDAFKOSPNSEH---- 1012
 Db 2152 AACLSG-----FAFSPFNKTYLKFVSDYGR---NGFOA-----VCOHRLMQ 2190
 QY 1013 ---GHSQORTSGIRTSDDSEEDPYMNDTVPTSPADSTVLLAPSVODSG-----S 1061
 Db 2191 ANGRADQOQSSGAAR-----TLPTOLENTATYKASMLADVCGCRNGDOFSCP 2237

QY 1062 LHNSSGSESTYCMQAGDLPSDG-----DYDQDDYED-- 1097
 Db 2238 VHSWSTGEEV-----AGDILHRLGLADGMRGVTAMKNGVQWAEIAGHDVLDLVSDLE 2291
 QY 1098 -----GATSSSVTFNSYSGOSPDYRCV----- 1124
 Db 2292 LIRDPPOKSYRIVTEBGRPALSRRGPKVFGNS-----WDSDEWSTPROPOEHMKVLDS 2347
 QY 1125 ---GTYNISGARYFSESGNOSFEDSEED-----FDSRFDTDDELSTYRD 1166
 Db 2348 DQYSHNDQGT-NGETEAQRGTATQESDLSGEPAVPHKGLDCLYLSLF--DPVLSYDGA 2404
 QY 1167 SVYSCVTLPEYHSLYMKGLM-----NSWRKNCVLK---DETFLWFSKOBA 1212
 Db 2405 DLEKPTALAY-----RMKGGGQPGGSSSGTEDTFRPRPEPKPIGDLASTL-----A 2452
 QY 1213 LKQGLHKKGGSSSTLSRRMKKRFVLROSL-----MYFENSEEKLKGTVEVRT-- 1264
 Db 2453 LQOAFIHKQ---AVLLARGMTLQATLQOQPLSALBSLPAEKPPABEAOFT-SVGTGP 2507
 QY 1265 -AKEIIDNTT-----KENGIDIIMADRTPHLLESPEDEASQ-----WF 1301
 Db 2508 PAKPVLLKATPKPLAPLAKAPRLPIKPVAPVLAQQA-----SPETTSPELVRV 2561
 QY 1302 SVLSOVH-ASTDOEIOEM-----HDE----- 1321
 Db 2562 STLNSEHFPPOPTQIKNTVROYQOPFGRGPRALKRDKGKVFMKRDPDEALMILKQOM 2621
 QY 1322 ---QANPONAVGTLVDGLDSVCAQSDP-----RPN 1351
 Db 2622 THLAAFGTOVSREAVALVKPTVSAPRPSMAPTSLPERSLBPBELTQTRLHLRLNPF 2681
 QY 1352 F-----VIITANVLH-----C----- 1363
 Db 2682 YQOQAPKIFLRKEVFYFKOSYSHNPVQDLDFRLIHLDTLSEALRISDEBRLMKALF 2741
 QY 1364 ---NADTPBEM-----HMITLLQ 1379
 Db 2742 AQONQDLOKPLVTEBSKRAVAVSTADTWEVFSRIFPATGSGVGVQLLAWSHVGIKLLR 2801
 QY 1380 RSKGDTVRGGEFIRGMLHKEVKNSSPKMSSLKUKKRFVL--THNSLDYKSSSK----- 1433
 Db 2802 MYKGGQEZAGQOLRVRAVSPADI-----LVTPMSQMLFNLASKEVILF 2847
 QY 1434 --NALKGLTV-----LNSLCSVVPDEKIF-----KETGY- 1462
 Db 2848 SARBAQVNTIVDDFLELKQSDVYAVARNFLPDEBPALAFHKGDIIHLQLEPPRVGS 2907
 QY 1463 -----MNV-TVYGRKHCYRLYTKULNEAT----- 1485
 Db 2908 AGCVRRRVVYLEELRRRGPDPFGWRFGTIHGR--VGRPSEILVOPAAAPDFQLPTEBGR 2965
 QY 1486 -RMSVIGNVDTKAPIDT-----PTQOLIODIKENCL---NSDVE--QIYKRN- 1530
 Db 2966 GRAAAVALAAVSAALAOEVRRRRREGPRVARSADGBEALALPRTYMLPEAKKTRIDPOR 3025
 QY 1531 -----ILRYTHNLHSLPLRPLRYGDIINLNLKDGTYTLQDEAI 1569
 Db 3026 RPQDGLRAKSKPRRESRTLEMLCTKTPLOESLIEU--SDSSLSKATDWFL-----AV 3078
 QY 1570 KIFNSLQOLESMSDPIITIGLLOTGNDLRLRDELYQOLIKQNKVPHPSVGNLYSQ 1629
 Db 3079 MRFMDAPLKGQSD-LDVLGNLKLKCGHEWREDECYOVVQKQITDNTSKODSCORQMR 3137
 QY 1630 ILTCL-----SCTPLPSRGILKYLFKHLKRIE--QPGTEMBKALFTYESLKKT--K 1679
 Db 3138 LLYITAVHSCS-----EVLHPLHTRPLQDVSRTPGRLPQGIKAKCBQNLQTLTLPFG 3189
 QY 1680 CREFVSRDEIEALI-HRQEWTSYVYCHGGGSKTITINSHTAGVEVKLIRGLAM--ED 1736
 Db 3190 GRLEIPSSIELRAMLAGSSSKROFLPLRGGLEHNLKIKTCVVALDVBEICAEMLATPBE 3249
 QY 1737 SRNMFALFEYNG--HVDKAIESRTVADVILAKEK-----LAATSEVGLDLPWKFYKFL 1787

QY 6 TRVLRNGOHFPESTVNSCAE-----GIVFRDYGVFTYKOSTITTHOKVTAMHP 56
 DB 11 TRVWIPD-----PEVVRSAELTKDYKDGESLOLRLEDDTIIDYPRIDVQNNQVPLRNP 65
 QY 57 TNEBGVDMASTLELHGSIWNLFORV-KRNOIWTYGSILASVNDYQPIAGLYEPATM 115
 DB 66 DILVGEEDLTALSHLEBPVAVLHNLKVFLESNMHYTCGIVLVAINVEQLP-IYGQDVI 124
 QY 116 EORSRHLGELPHPIFALANECVACLKMRHNOICILKGSAGKTESTLILKFLSVIS 175
 DB 125 YASGQNMGMDDPITPAVABAYQOMARDEKNOIIVSGESGAKTVSAYVAMRYEFTVG 184
 QY 176 QOSLELSLKEKTSCEVERALIESSEPIMEAFGNKTVNNSSRFGKFOVOLI COGNIOG 235
 DB 185 GSAADNTIEEK-----VLASSPIMEAIGNAKTRINDNSRFGKIYIEIGFDKHYHIGA 237
 QY 236 RIYDCILSSONRVYRQNGERNYHIFYALLAGLEHEBEREFPYLSTPENHYLNOGCVED 295
 DB 238 NMRTYLL-EKSRVYFQADDERNHYHIFYOLCAASLPEFKELALTCADPFYTAHGGNTTI 296
 QY 296 KTISDOESFEVITAMVMOFSKEVREVSRLLAGILHGNIEF-ITAGAQVSKFA-- 352
 DB 297 EGVDDADPEKTRQALTLTGVRSHQISIFKIIASILHLSVEIQERDDSCSISPODE 356
 QY 353 -LGRSAELLGJDPOTLTALTORSMFLRGEIILTPLVQOAVDSRDSIAMAALYACCEWV 411
 DB 357 HLSNFCLLGIBHSQOMHMLCHRKLVTSSETYKTSLOQVAVARANALAGIYAOLFSW 416
 QY 412 IKKINSRIKON-EDFSGISGLIDFGFENFEVNHFEQFNINAYANEKLOEYENKAI FSLLEOL 470
 DB 417 VERINKALQTSLOKSHSPFIVGLDIYGFETFEINSEFOCINAYANEKLOOQFNSHVFLEOE 476
 QY 471 EYREGLVWEDIMIDWIDGECLEIEKKGLLALINESHPPOATDSTLEKLSOHANN 530
 DB 477 EYKGOIPLWIDPYDQPCIDILEAKGLIDLEDECKVPKGTDOMAOKLERHSNSO 536
 QY 531 FYVPRVAVNPGVKHAYGEVOYVRIELEKNRDTFADDLNLLRESRPFIDVLE--- 586
 DB 537 HFQKPRMSNTAFIVIHADKVEYLSDFLEKONDYVEEQINILKASKPLVADLRDE 596
 QY 587 -----EHVSSRNNODTLKCGSKHRPVSOSQFKVDSLHSLMATSSNPPF 632
 DB 597 DSVPATNTAKSRSSSKINVRSSRPLMKAPNKEHKSQYOFRT-SINLMEETINATTPHY 655
 QY 633 VRICIKPMQKPOPODAVVLNQLRYSGLMLETIRIKAGAVARPPODEFYKRYKVM--R 690
 DB 656 VRICIKPNDKLPHPFDKRAVVOQLRACGVLETRISAGYPSRWYTHDFPNRIRVLMKKR 715
 QY 691 NLALPEVDRGCTSLQLYDASNEWOLGKTVFLRESLEOKLEKREBEREESHAAVIRA 750
 DB 716 ELANTTDKKNICGSVLESILKDPDKQFGRTKIFPAGOVAYALEKLRADFRATIMICK 775
 QY 751 HVLGFLARKOYKRVLYCVVITQONVAFLLRRFLHLK--KAAIYQOKOLRGCIARVY- 807
 DB 776 TVRGMLOVRKYRRLRAATLTLQFCRGYLLARLLEHLRRTRAIIVQOKOYRMLKARRAVC 835
 QY 808 ----- 807
 DB 836 RVRRAVVIOSYTRGHVCTOKLPVULTEKATIIQYANGWARRHFORORDAIVIOCA 895
 QY 808 -ROLLAEKEBOEKKOQEEBKKKREEREREREREREAALRAQOEETRAKOELALOK 866
 DB 896 FRRLKARQALKALKEARSSEHLKRLNVGMENKVVOLOKRIIDONNEFTLSQLSAYVS 955
 QY 867 --SOKAEITRELEKOKENKOVETIIRLEKEIEDLOKMEQOELSTIASLOKLOERRDO 924
 DB 956 THAMEVERKKEKELARVOONOEADPSLOLEBYOULR--TELQVMSERVLJEHANHRENG 1013
 QY 925 ELUR-----LEEBACRAOEFLBSINDEIDECVRNTERLSGSEBSSELSAAGECKEN 980
 DB 1014 ELURKRAVDLEHNAALKDE-KETHLN-----HOIURKSSASSOSSVEENLILKEEJEENS 1068

QY 981 --FNFSQPYR-EEVDEGFEADDDAFKOSPNESEHG--HSDORTSGIRTSDDSSSEDPYM 1035
 DB 1069 RYONLVKEYSOLQORENENRDEQOTPGHNRKNNSGSSLESDDNYSISITSEIGDEIDALQ 1128
 QY 1036 NDTVPTSPSA-DSTYLL-----APSVODSSSLH-----NSSGSESTYCMPONAGDLP 1083
 DB 1129 QVEEIGIEKAAADMTVFLQLQVRVLEBOERKLDVQLEKEQODSKQVQVEOQNNGLDVD 1188
 QY 1084 PGDDVYDQDDVEDGATSGSSVTFSSNSYGQMSDPYRCSVGTYNSSGAYRPSSEGAQS 1143
 DB 1189 QADADIYNLSKOE--LESENKLLKNDLENK-----AVADOAMQDN 1229
 QY 1144 FEDSEDPDSRF-----DTDELSTYRDSVSVSVTLPYFHSFLYMGGLMSMKRWCVL 1198
 DB 1230 SHSSPDSYSLNLNQKLANELEVREKEV-----LIRTOGMANDQRLSLGK 1277
 QY 1199 KOTFL-----WFRSKOELKOGMHLKKGSSSTLSR-RNKKRPFVLROSLMTFENDS 1252
 DB 1278 NMEPNINARTSWPNEKHWODADAEVYHGVQVTSQTEDMG-----YLNBDG 1325
 QY 1253 E-----EKLGVETRTAKEIIDNTTKENGIDIIIMADRTFHLIASPEDASQMF----- 1301
 DB 1326 ELGLAYQGLQVAYARLEAQLQONLKHREVEHLQAO--VEAMKEBMDQOQOTFCOTLL 1383
 QY 1302 SVLSQVHASTDOEIOEMHDEQANPONAVGTLD-----VGLIDSVCASDS 1345
 DB 1384 SPEAOVEFQOQEIIRLTJENENDPFKELEVKEKNEKKLKKOLKIYMKVQODEAQAOLAO 1443
 QY 1346 PRPNRSFVITANRVLHCADTPREEMHWIT--LLOSKGDTREVGGEFIVRGUL--HK 1400
 DB 1444 SRR-----RHHELTPQVTVORERXD-----FOGMLEYHK 1472
 QY 1401 E-----VKN-----SPKMS-----LKLKRFVLTNHSIDYKSEKXALGLTVNS 1445
 DB 1473 EDEBALLIRLVLDLQPMLSGTVPLPAYIILMCIRHAD--YTMDLKVHSLSTING 1529
 QY 1446 LCSVPRPEKIPKETGYMNVVYGRKCYRLYTKLLNATRWSVYIQNVTDKABIDPT 1505
 DB 1530 IKKVLKHNEDPEMTSFMLSNTRCLHCLKQYSGDEGFWTO-NTAKON----- 1576
 QY 1506 QOLIODIKENCL--NDDVBOIYKRNPIRLYTHPLHSLPLRPYGDINLNLKDGYYTL 1564
 DB 1577 -----EHCKNDLDE--YRQ-----VL 1592
 QY 1565 ODEAIKIFNSLOOL-ESMSDPIPI-----IOGILQTHDLRP----- 1600
 DB 1593 SDSLQIYQOLIKIABGLQPMIVASAMLENISQLSGVRPFGYKXSSMWDSNYSYCL 1652
 QY 1601 -----LRDELXQOLI KQTNKVPHPSVGNLYSMOILTLCTFLPSR 1642
 DB 1653 EAIIRQMNPFHTVLCDOGLDPEIILQVFKQLPYMINAVTLNLL-----LLRKDACSM--ST 1706
 QY 1643 GILKLVKHLKRIKQPFQTEMEKY-ALFYTESL-----KTKCREVPSRDELE 1691
 DB 1707 GM--OLARNISQLEBMLGKMLQOSGAQVOTMPEPLQAAQLOLKKT-----QEDAE 1756
 QY 1692 AL-----IHRQEMSTYVCH--CGSGCKITIN-SHTTAGEVVEKILRGLAMEDSRNMF 1741
 DB 1757 AICSLCTSLSTQOIKVILNLYTPPLNGFEERVTVSFRTIQAOLOQERSDPOOLILDSKMF 1816
 QY 1742 -ALFEYN 1747
 DB 1817 PVLFPFN 1823

RESULT 14
 A26655
 myosin heavy chain [similarity] - alime mold (Dictyostelium discoideum)
 N:Contains: myosin ATPase (EC 3.6.4.1)
 C:Species: Dictyostelium discoideum
 C>Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 19-Apr-2002
 C:Accession: A26655, A24728, S00250
 R:Warrick, H.M.; De Lozanne, A.; Leinwand, L.A.; Spudich, J.A.

QY 70 ELHGSIWYNUFORYKXNOIWTYIGSILASVNPYOPAGLBYEPATMEQYSRHGLBPH 129
 DB 156 ELTEGNNLKXKJHFLQOIKYTYAGSILVAINPFR-PLPIYKPYVMYENQOGLXEPH 214
 QY 130 IFAIANCYRCMLKRNOCILIKGESGAKTBTSTKILKFLVSTISQOSLELSKXTSC 189
 DB 215 VFLADAVAYTYMKRKNOCIVYPGESGKTOSTNLHLCTALSG-----KGASG 267
 QY 190 VERALESSEPRIMEAFNAKTIVNNNSRFGKFOVLANICQKNIQGGIVDCILSQRNV 249
 DB 268 VERITILGACVLEHFGAKTAHNNNSRFGKFIQVSTLESGYRGA-VAKTILLESKRLV 326
 QY 250 RQNGERNYHIFVALLAGLEHEEREFPYSTPENYHYNOSGVEDKTISSDOE---SFR 305
 DB 327 SOEDERNYHIFVYLLGVSEERQEFOLKOPEDYFANQ-----HNLKIDGDLKHDFE 382
 QY 306 EVITAMVMPKSEBNAEVSRLAGILHLNIEITNG-----GAQVSEFTALGRSALL 360
 DB 383 RLKQAMEMVGFPLATKQIFAVALSAILYGNVYKRAATGREBGLGVPEVLDTLSQLL 442
 QY 361 GLDPTOLDALJORSFMRGEEILTPANVOAVDSRSLAMALYACCFEVIKINSRIK 420
 DB 443 KVRKEILVEVLTAKKYTVNDKILPYSLEAITARDMAKSLSYALFDMIVLRINHAL 502
 QY 421 GNEDFK-----SIGLIDIFGFENFEVNHFEQFNINANEKLOEYFNKAI FSLDLEYSR 474
 DB 503 NKDVEBAVAGSLSIGVADIFGFEDFERNSFEQFCINANBOLQYFNOHIFKLEOEYOG 562
 QY 475 EGLVMEIDWIDNGECIDLIEK-LGILLALINESHPQATDSTLEKHSQANNHXY 533
 DB 563 EGTWNIHIGYTDVNGCIHLISKKPTGLFYLLDESNPHATISQTLAKFKQOEDNRYFL 622
 QY 534 KPRVAVNPFVCYVAGEVOYDVRGILEKNRDTFRDDLNLRESRFDYDLF----- 586
 DB 623 GTVMEBAFIIOHFGAKVYKQIKDFREKNMDYMRPDLVALLRGSDDSYABELGMPVAV 682
 QY 587 -----BH----- 588
 DB 683 FRMAVLRAIRAMAVLREAGRLERAERAKAGMSSPQAGHPRELPAGASTPSEKLYRL 742
 QY 589 ----- 588
 DB 743 HNOIKSICGLPMOGEDPRSLSLOSLRQPRAFILKSKGKIKOKOII PKNLDSKSKL 802
 QY 589 -----VSSRNNODTLKSGSKHRRPTVSQOFVDSLHSLMATLSSSNPFVVCIPKNOQKMP 644
 DB 803 ISMTLDRHTTKSLHLHKKKKPSISAGPOT-SLNTLLEALGKAEPFRICISMAEKE 861
 QY 645 DQFDQAVNLQRLYSGMLETVRIRKAGYAVRRPDPYKRYKYLMBNLALPEDV---RGK 701
 DB 862 LCFDDELIVLOQLRYTGMLETVRIRRSYSAKYTFODFTBOFVL-----LPKDAQPCREV 916
 QY 702 CTSLLOLYDASNEMLGKTQVFLRBSLEKLEKREBEVSHAAMTIRAHVGLFARKOY 761
 DB 917 ISLLEBMKIDKNYOGKTKVFLKETERQALQETLHREVVARKILLLOSFRMVLRRHF 976
 QY 762 RKLVCVILQKRYRAFLRRRFLHLKKAIVFOKOLRGOLARVYR-----QLA 812
 DB 977 LOKKRAAVTIQACMRGRVRA-LERTQAAVYLQAAWGYWOKLIRHQKOSTIRLOQSLC 1035
 QY 813 EKREOEKKKQEEBEKKREERER-ERERREABLRQOEEBTRKQOELEALQSKOKA 871
 DB 1036 RGHLOKRSFQOMISEKOKAEKERALEABRAGAGGQOAGGQVABOGEPRADGG 1095
 QY 872 ELTRELEKQENQVEILRLKEIEDLOMKEQOELSTEASLOQLQERRODELARLEE 931
 DB 1096 HLASBEVQSDSPLEHSSPEKEAPSPEKTLPPOKTVAESHEKVPSSSEKESRR--- 1152
 QY 932 EACRAQOEFLSINFP--EIDECVRNIERSLSGGS--BFSSELAESACEEKPNFNSQPYR 988
 DB 1153 -----QRGLEHVKFQNKHIOCKE--ESALREPSRRVTOEOGVSLBEDK-----K 1195
 QY 989 EEEVDEGF-----EADDAFK--DSPNPSEHGHSDORTSGIRTSDDSEEDPYMNDIVV 1040

DB 1196 ESRDEBTLLVVEATEANTSOQOPTEQPQAMAVGVSEETE--KTLPSGS---PRGQLER 1250
 QY 1041 PPSPADSTVLLAPSVQDGSGLHSSSGSESTYCMQNGDLP-SPDGDVDYQDDYEDG- 1098
 DB 1251 PTLSDLSRV---SPAPGAPETPEDESKPGSPROVEKDSFGSGTOIOR--YLDAE 1304
 QY 1099 -----AITSGGSVTFPSNYSQOWSPDYRCVGTNNSGAVRFSSEGAQSPEDSE--- 1148
 DB 1305 RLASAVELMRGKVLVAASPSAMLSQSLDS--DRHATGALALTPREBRTSTSDVSLK 1363
 QY 1149 -----EDPDRPDTDELSTYRDSYSCVTLPEFHSFLYMKGLMNSWGRKRCVLKDET 1203
 DB 1364 LPSLAKAPPAEATTGERSAKKPAVOK----- 1390
 QY 1204 LMFRSQOBLKQGMHLKGGSGSTL-----SRNKKRMFVLROSKLMYFEND 1251
 DB 1391 -----KRGDASSLPDAGLSFGSOVDSKSTRK--LFLHKT-----D 1426
 QY 1252 SEEKLGVEVRTAKE---IIDNTKENGIDITIMADRTFHIAE---SPEDASQMFVYL 1304
 DB 1427 KYYSLEGAEELENNVAGHVLEATTMKKGLBAPSGQHRHNAAGEKTEPGKGGKNNRV 1486
 QY 1305 SOVHASTDQEIQEMDEQANPONAVGTLVDGLIDSVCASDSPDRNSFVITANRVLHN 1364
 DB 1487 KJGKLTSEKMBESVFRQITNANBLKYLDEFLNKINILRSQKTEPESLFIATEKFRSN 1546
 QY 1365 ADT-----PREMHHTL-----QREKGRVAGQEF-----YVGMHLK 1400
 DB 1547 IKTMYSVPNGKIHVGKDMENYQIVNSNLATERQOKTNLVLNLFQSLDDEFTGTYKN 1606
 QY 1401 E---VNSPKNSSLLKMKMFVLTHNS---LDYKSSSEKNALK-----LGTVLNLSLC 1447
 DB 1607 DEEPYKQS-KAQKKRQOERAVQEHNGHVFPASVYQISQSGCGLSYIMWDKALLCSVC 1665
 QY 1448 SVPPDEKIFKETGYMNTVYGRKHCYRLYKLNLEATRWSSVIONVDTKAPIDTPPOQ 1507
 DB 1666 KMTCHKKCVHKIQSHCSYT-YGRK-----GEPGAPGHFGVCVDSLTSDKASVPVLEK 1718
 QY 1508 LQODIKENCLNSDVVEOYKRN-----PLRYHHPHLS----- 1541
 DB 1719 LLEHVEHNGL---YBGLYRKSGAANRTRELQALQTPPAKMLKEMPIHALTYGLKOWL 1775
 QY 1542 -----PLPFL--PYGININLNLKDKGYTLLODEAIKIFNSLOQLESMSDPIPIIOGILQYG 1595
 DB 1776 RELPEPLMTFPAQYDFL-----RAVELPKQOQLAM---YAVLHMLPEAN 1818
 QY 1596 HDL--RPLRDELYCOLIKQTKVPHPSGVNLYSQOILTC--LSTFLPSRGILKYLFH 1651
 DB 1819 HNSLERLIFHLVAKVALLBVDVRM--SPGALAIIFAPCLRCPDNSDPLTSMKDVLEKITTVCV 1877
 QY 1652 LKRIEOPRGTEMEKALFTYESLKKTKCRFVPSRDEIBALIHQWMTSTVYHGGGSC 1711
 DB 1878 EKLITEQ---WKRYKV-KNEEISQLEABASIAR-RLSLRQWANKSPKTRPAGGAG 1930
 QY 1712 KITINSHTTAGEVEKIJRG 1731
 DB 1931 RLITTSRVSPSPSTNNALG 1950

Search completed: July 14, 2003, 18:24:10
 Job time : 68.7074 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:46 ; Search time 21.3952 Seconds

(without alignments)
3987.653 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730

Sequence: 1 FCIQGRVWLRENQHPST.....MIVKKYSTRSASSQSSR 2057

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10626	99.0	2058	MY10_HUMAN	Q9h667 homo sapien
2	9961	92.8	2052	MY10_BOVIN	P79114 bos taurus
3	1840.5	17.2	2215	MY7A_MOUSE	P97479 mus musculu
4	1831	17.1	2215	MY7A_HUMAN	Q13402 homo sapien
5	1593	14.8	3511	MY15_MOUSE	Q9qz24 mus musculu
6	1589.5	14.8	3530	MY15_HUMAN	Q9ukn7 homo sapien
7	1504.5	14.0	1846	MY5B_RAT	P70569 rattus norv
8	1500	14.0	2116	MY5B_DICDI	P08799 dictyostell
9	1492	13.9	2158	MY9B_HUMAN	Q13459 homo sapien
10	1490.5	13.9	2245	MY9B_DICDI	P54697 dictyostell
11	1471	13.7	1980	MY9B_RAT	Q63358 rattus norv
12	1466	13.7	1855	MY5A_HUMAN	Q9y411 homo sapien
13	1448.5	13.5	1742	MY5C_HUMAN	Q9nqf4 homo sapien
14	1444.5	13.5	1828	MY5A_RAT	Q9qy43 rattus norv
15	1442	13.4	1574	MY92_YEAST	P19524 saccharomyc
16	1440	13.4	2114	MY92_MOUSE	Q9qy06 mus musculu
17	1434	13.4	1853	MY5A_MOUSE	Q99104 mus musculu
18	1432	13.3	1829	MY5A_CHICK	Q02440 gallus gall
19	1393.5	13.0	1262	MY0E_HUMAN	Q9un54 homo sapien
20	1373.5	12.8	1470	MY5N_ACACA	P05559 acanthamoeb
21	1362.5	12.7	1501	MY5N_YEAST	P32492 saccharomyc
22	1344	12.5	1978	MY5B_CHICK	P10587 gallus gall
23	1341.5	12.5	1972	MY5B_RABIT	P35748 cryotolagus
24	1341	12.5	1938	MY5B_CHICK	P13538 gallus gall
25	1333	12.4	1972	MY5B_HUMAN	P35749 homo sapien
26	1331.5	12.4	1940	MY5B_CHICK	P02565 gallus gall
27	1323	12.3	2017	MY5N_DROME	Q99323 drosophila
28	1322.5	12.3	1928	MY5N_YEAST	P08864 saccharomyc
29	1320.5	12.3	1972	MY5B_MOUSE	Q08638 mus musculu
30	1320.5	12.3	1976	MY5B_HUMAN	P35580 homo sapien
31	1319.5	12.3	1976	MY5A_BOVIN	Q27991 bos taurus
32	1317.5	12.3	1976	MY5A_RAT	Q9j100 rattus norv
33	1314	12.2	1941	MY5B_HUMAN	Q9ukx2 homo sapien

34	1313.5	12.2	1938	MY5A_BOVIN	P24733 aequipecten
35	1311	12.2	1937	MY5B_HUMAN	P13535 homo sapien
36	1306.5	12.2	1969	MY5A_CAEL	P12844 caenorhabd
37	1297.5	12.1	1960	MY5B_HUMAN	P35579 homo sapien
38	1294	12.1	1935	MY5B_HUMAN	P12883 homo sapien
39	1288	12.0	1938	MY5D_CAEL	P02567 caenorhabd
40	1287.5	12.0	1935	MY5D_DICDI	P34109 dictyostell
41	1286.5	12.0	1935	MY5D_PIG	P79293 sus scrofa
42	1286	12.0	1966	MY5B_CAEL	P02566 caenorhabd
43	1285	12.0	1938	MY5B_MOUSE	Q02566 mus musculu
44	1279	11.9	1939	MY5B_HUMAN	P12882 homo sapien
45	1278	11.9	1959	MY5B_CHICK	P14105 gallus gall

ALIGNMENTS

RESULT 1

MY10_HUMAN STANDARD; PRT; 2058 AA.

AC Q9HD67; Q9NYM7; Q9P110; Q9UHF6;

DT 15-JUN-2002 (Rel. 41, Created)

DT 15-JUN-2002 (Rel. 41, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Myosin X.

GN MYO10.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Carnivora; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20442246; PubMed=10984435;

RA Berg J.S., Deffler B.H., Pennisi C.M., Corey D.P., Cheney R.E.

RT "Myosin-X, a novel myosin with pleckstrin homology domains, associates

RL with regions of dynamic actin."

RT J. Cell Sci. 113:3439-3451(2000).

RN [2]

RP SEQUENCE FROM N.A.

RA Rogers M.S., Strehler E.E.

RT "Identification of myosin X as a specific binding partner for the

RT tumor suppressive calmodulin-like protein."

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA TISSUE=Colorectal adenocarcinoma;

RT "Cloning of human myosin X."

RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE OF 347-495 FROM N.A.

RA TISSUE=Skeletal muscle;

RT MEDLINE=20079149; PubMed=10610710;

RA Rojce K., Serrano de la Pena L., Gallardo T., Simons A., Myce K.,

RA McGrath R.J., Considine E., Vasko A.J., Peterson E., Grady D., Cox R.,

RA Andrew L.J., Lovett M., Overhauser J., Williams C.J.

RT "Physical map and characterization of transcripts in the candidate

RT interval for familial chondroblastoma at chromosome 5p15.1."

RL Genomics 62:177-183(1999).

CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE

CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.

CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS

CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY

CC SIMILARITY). PLAYS A ROLE IN REGIONS OF DYNAMIC ACTIN.

CC -1- SIMILARITY: CONTRAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC -1- SIMILARITY: CONTRAINS 2 PH DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AP247457; AAF68025.2; -
DR EMBL; AF234532; AAF37875.1; -
DR EMBL; AF132021; AAF36524.1; -
DR EMBL; AF132022; AAF36525.1; -
DR EMBL; AF184153; AAF17363.1; -
DR HSSP; P08799; 1MND.
DR Genew; HGNC; 7593; MYO10.
DR MIM; 601481; -
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR000857; MYTH4.
DR InterPro; IPR001849; PH.
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00063; myosin_head; 2.
DR Pfam; PF00169; PH; 3.
DR Pfam; PF00612; IQ; 3.
DR Pfam; PF00784; MYTH4; 2.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00295; B41; 1.
DR SMART; SM00015; IQ; 3.
DR SMART; SM00242; MYSC; 1.
DR SMART; SM00139; MYTH4; 1.
DR SMART; SM00233; PH; 2.
DR PROSITE; PS00660; BAND_41.1; FALSE_NEG.
DR PROSITE; PS00661; BAND_41.2; FALSE_NEG.
DR PROSITE; PS50057; BAND_41.3; 1.
DR PROSITE; PS50096; IQ; 2.
DR PROSITE; PS50003; PH_DOMAIN; 2.
DR Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat.
KM Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat.
FT DOMAIN 1 727 HEAD OR MOTOR DOMAIN.
FT DOMAIN 742 763 IQ 1.
FT DOMAIN 764 787 IQ 2.
FT DOMAIN 788 817 IQ 3.
FT DOMAIN 1212 1310 PH 1.
FT DOMAIN 1392 1497 PH 2.
FT DOMAIN 1697 1981 BAND 4.1-LIKE.
FT DOMAIN 799 943 COILED COIL (POTENTIAL).
FT CONFILCT 98 98 S -> P (IN REF. 3).
FT CONFILCT 148 148 H -> Y (IN REF. 2).
FT CONFILCT 256 256 G -> W (IN REF. 1).
FT CONFILCT 324 324 W -> R (IN REF. 2).
FT CONFILCT 1186 1186 G -> C (IN REF. 3).
FT CONFILCT 1663 1663 T -> S (IN REF. 3).
SQ SEQUENCE 2058 AA; 237389 MW; E4E19BF92D87F3AC CRC64;

Query Match 99.0%; Score 10626; DB 1; Length 2058;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2043; Conservative 4; Mismatches 8; Indels 2; Gaps 2;

QY 1 FCLQGRVWLRENGQHPSTVNSCAEGIVFRTDYQVFTYKOSTITTHQKVTAMHPNNE 60
DB 4 FTEGGRVWLRENGQHPSTVNSCAEGIVFRTDYQVFTYKOSTITTHQKVTAMHPNNE 63
QY 61 GUDNMASTLTHGSGIMYNLFQRYKRNQITWYIGSLASNPPOPIAGLYEPTMYOYR 120
DB 64 GUDNMASTLTHGSGIMYNLFQRYKRNQITWYIGSLASNPPOPIAGLYEPTMYOYR 123
QY 121 RHILGELPHIFAIANECYRLMRHNOCLIKSGSAGKTESTKILKFLSVISQOSLE 180
DB 124 RHILGELPHIFAIANECYRLMRHNOCLIKSGSAGKTESTKILKFLSVISQOSLE 183
QY 181 LSLIKETSCVERAILESSPIMEAFGNAKTYNNNSRFGKTYOLNI CQKGNIGGRIIVDC 240
DB 184 LSLIKETSCVERAILESSPIMEAFGNAKTYNNNSRFGKTYOLNI CQKGNIGGRIIVDC 243
QY 241 ILSNORVVRONGENNYHFFYALLAGLEHEEREETLSPENHYHLNOSGVEDKTID 300
DB 244 LLSNORVVRONGENNYHFFYALLAGLEHEEREETLSPENHYHLNOSGVEDKTID 302

QY 301 QESFREVITAMVMOFSKEEVEVSRLLAGILHNLNIEITTAGAQSFKTAGSABLL 360
DB 303 QESFREVITAMVMOFSKEEVEVSRLLAGILHNLNIEITTAGAQSFKTAGSABLL 362
QY 361 GLDPTQLTDALTQSRMFLRGEIILPLVYQQA VDSRDSLAAALYACCEBWTIKINSRIK 420
DB 363 GLDPTQLTDALTQSRMFLRGEIILPLVYQQA VDSRDSLAAALYACCEBWTIKINSRIK 422
QY 421 GNEDFKSIGLIDIPGFENPEVNHFPQFNINYNENKLOEYFNHGISFLQLEYSRGLWE 480
DB 423 GNEDFKSIGLIDIPGFENPEVNHFPQFNINYNENKLOEYFNHGISFLQLEYSRGLWE 482
QY 481 DIDWIDNGECDLIEKKLGLLALINEESHFPQATSTLLEKLSQHANNH FYKBRVAVN 540
DB 483 DIDWIDNGECDLIEKKLGLLALINEESHFPQATSTLLEKLSQHANNH FYKBRVAVN 542
QY 541 NFGVXHYAGEVOYDVRCILIEKRDTPRDDLNLBESRPFYDL.FEHVSSRNNODTLKC 600
DB 543 NFGVXHYAGEVOYDVRCILIEKRDTPRDDLNLBESRPFYDL.FEHVSSRNNODTLKC 602
QY 601 GSKHRRPTVSSQFKVDSLSHMAITSSSNPFVRCIKPMQKMPQDFQAVVNLQRLYSG 660
DB 603 GSKHRRPTVSSQFKVDSLSHMAITSSSNPFVRCIKPMQKMPQDFQAVVNLQRLYSG 661
QY 661 MLETVRIRKAGYAVRPFQDFYKRYKVMRNALPEVNGKCTSLLOLYDASNSEMOLGK 720
DB 663 MLETVRIRKAGYAVRPFQDFYKRYKVMRNALPEVNGKCTSLLOLYDASNSEMOLGK 721
QY 721 TVYFLRESLEQLEKREBEVSHAAVIRAHVYGLTAKQRYKVL CYVITIQNTRAFLL 780
DB 722 TVYFLRESLEQLEKREBEVSHAAVIRAHVYGLTAKQRYKVL CYVITIQNTRAFLL 781
QY 781 RRRFLHLKKAALVFOKQRLRGQIARVRYQLAEKKEOEKKOEKKEEKKREERERER 840
DB 782 RRRFLHLKKAALVFOKQRLRGQIARVRYQLAEKKEOEKKOEKKEEKKREERERER 841
QY 841 ERREBELRAQOEESTRKOQLEBALOKSQKEALATRELEKOKENKOVEILRLKEIEDLO 900
DB 842 ERREBELRAQOEESTRKOQLEBALOKSQKEALATRELEKOKENKOVEILRLKEIEDLO 901
QY 901 RKKEQOEISLTASIQLOERRDQELRLLEBACPAQOFLSLNPFDEIDECVRNIERSL 960
DB 902 RKKEQOEISLTASIQLOERRDQELRLLEBACPAQOFLSLNPFDEIDECVRNIERSL 961
QY 961 SGSGSFSSSLASACEBKPNFNSQPYPEBEVDEGEFADDAFKQSPNPSBEHSGSDQRTS 1020
DB 962 SGSGSFSSSLASACEBKPNFNSQPYPEBEVDEGEFADDAFKQSPNPSBEHSGSDQRTS 1021
QY 1021 GIRTSDSSSEBDPYNNDTVPTSPASDSTVLAPSVODSGSLHNSSGESTYCMQONAGD 1080
DB 1022 GIRTSDSSSEBDPYNNDTVPTSPASDSTVLAPSVODSGSLHNSSGESTYCMQONAGD 1081
QY 1081 LPSPPGDYDYDDDYEDGALITSGSVTFNSYSGQMSDPYRCSVGTYNSSGAYRPSBEA 1140
DB 1082 LPSPPGDYDYDDDYEDGALITSGSVTFNSYSGQMSDPYRCSVGTYNSSGAYRPSBEA 1141
QY 1141 GSSPFDSEEDPDRDPTDELSTYRDSVYSCVTLFPHSFLYMKGGLNMSWGRKRCVLYD 1200
DB 1142 GSSPFDSEEDPDRDPTDELSTYRDSVYSCVTLFPHSFLYMKGGLNMSWGRKRCVLYD 1201
QY 1201 EFTLWFRSKOBLKQWMLHKKGGSSSTLSRRMKRWVFLRQSKLMEFENDSEBKLKGV 1260
DB 1202 EFTLWFRSKOBLKQWMLHKKGGSSSTLSRRMKRWVFLRQSKLMEFENDSEBKLKGV 1261
QY 1261 EYRTAKEIINDTTKENGIDIIIMADRTFHLIASEPDAQSWFVLSQVASTQOEIOENHD 1320
DB 1262 EYRTAKEIINDTTKENGIDIIIMADRTFHLIASEPDAQSWFVLSQVASTQOEIOENHD 1321
QY 1321 EOPANQAVGTLDVGLIDSVCSADSPDRPNSFVITITANRVLHCNADTPEBMMHWITLQOR 1380
DB 1322 EOPANQAVGTLDVGLIDSVCSADSPDRPNSFVITITANRVLHCNADTPEBMMHWITLQOR 1381

QY 1381 SKGDTREVGQEPITVGMHLKEVNSPKNSLTKKRWFLVTHNSLDYKSSSEKALKIGT 1440
 DB 1382 SKGDTREVGQEPITVGMHLKEVNSPKNSLTKKRWFLVTHNSLDYKSSSEKALKIGT 1441
 QY 1441 LVNLSCSVVPDEKIPKETGYMNTVYGRKHCYRLYTELNEARWSSVIONMTDTPR 1500
 DB 1442 LVNLSCSVVPDEKIPKETGYMNTVYGRKHCYRLYTELNEARWSSVIONMTDTPR 1501
 QY 1501 IDPTQOQLIQDIKENCNSDVVEQIYKRNPIRLYTHPHLSPLPLPYGDINMLNKDGG 1560
 DB 1502 IDPTQOQLIQDIKENCNSDVVEQIYKRNPIRLYTHPHLSPLPLPYGDINMLNKDGG 1561
 QY 1561 YTTLODEAIKIFNSIQLESMSDPIPIIGIILQTHDLPLRDELYCQILKQTNKVPHG 1620
 DB 1562 YTTLODEAIKIFNSIQLESMSDPIPIIGIILQTHDLPLRDELYCQILKQTNKVPHG 1621
 QY 1621 SVGNLYSMOILTCLSTPLPSRGLKYLKFKLRIRBOCPGTEMEKALFTYESLKKTC 1680
 DB 1622 SVGNLYSMOILTCLSTPLPSRGLKYLKFKLRIRBOCPGTEMEKALFTYESLKKTC 1681
 QY 1681 REFVPSRDEIALIRHOEWSTVYCHGGSCKITINSHTAGEVVEKLRGLAMEDSRM 1740
 DB 1682 REFVPSRDEIALIRHOEWSTVYCHGGSCKITINSHTAGEVVEKLRGLAMEDSRM 1741
 QY 1741 FALFENYGHVDAKISRTVAVDLAKFEKLAATSEVGDJPMKFPYKLYCFDLDTPNVKDS 1800
 DB 1742 FALFENYGHVDAKISRTVAVDLAKFEKLAATSEVGDJPMKFPYKLYCFDLDTPNVKDS 1801
 QY 1801 VEFAPMFQAHAVYHGHHPAPEBNQVLAARLQLOLDYTLHAIPLEEVYSIORLK 1860
 DB 1802 VEFAPMFQAHAVYHGHHPAPEBNQVLAARLQLOLDYTLHAIPLEEVYSIORLK 1861
 QY 1861 ARISQSTKFTCEERLEKERTSFLBETLARSFRITSVVRQKVEBOMLMWKEEVSAR 1920
 DB 1862 ARISQSTKFTCEERLEKERTSFLBETLARSFRITSVVRQKVEBOMLMWKEEVSAR 1921
 QY 1921 ASIIDKRRKFGQMNQEQAMAKYMALIKENPGYSTLFEVECKEGFPQELMGVSADAVS 1980
 DB 1922 ASIIDKRRKFGQMNQEQAMAKYMALIKENPGYSTLFEVECKEGFPQELMGVSADAVS 1981
 QY 1981 VYKREGRPLEVFOYEHILISFGAPLANTYKIYVDEBELFETSEVVDVAKMKATYSMTV 2040
 DB 1982 VYKREGRPLEVFOYEHILISFGAPLANTYKIYVDEBELFETSEVVDVAKMKATYSMTV 2041
 QY 2041 KRYSTTSRASSOGSSR 2057
 DB 2042 KRYSTTSRASSOGSSR 2058
 RESULT 2
 MY10_BOVIN STANDARD; PRT; 2052 AA.
 AC P79114;
 DT 15-JUN-2002 (rel. 41, Created)
 DT 15-JUN-2002 (rel. 41, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE Myosin X.
 GN MYO10.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=heart;
 RX MEDLINE=20442246; PubMed=10984435;
 RA Berg J.S., Dettler B.H., Pennisi C.M., Corey D.P., Cheney R.E.;
 RT "Myosin-X, a novel myosin with pleckstrin homology domains, associates
 with regions of dynamic actin".
 RU J. Cell Sci. 113:3439-3451(2000).
 CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
 ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.

CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
 CC SIMILARITY). PLAYS A ROLE IN REGIONS OF DYNAMIC ACTIN.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 IQ DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U55042; AAB39486.1; -.
 CC HSSP: P08799; IAMD.
 CC InterPro: IPR000299; Band_4.1.
 CC InterPro: IPR000048; IQ_region.
 CC InterPro: IPR000857; MYTH4.
 CC InterPro: IPR001849; PH.
 CC InterPro: IPR001609; myosin_head.
 CC Pfam: PF00063; myosin_head; 1.
 CC Pfam: PF00169; PH; 3.
 CC Pfam: PF00612; IQ; 3.
 CC Pfam: PF00784; MYTH4; 1.
 CC PRINTS: PR00193; MYOSINHEAVY.
 CC ProDom: PD000355; myosin_head; 1.
 CC SMART: SM00295; B41; 1.
 CC SMART: SM00015; IQ; 3.
 CC SMART: SM00242; MYSC; 1.
 CC SMART: SM00139; MYTH4; 1.
 CC SMART: SM00233; PH; 2.
 CC PROSITE: PS00660; BAND_4.1; FALSE_NEG.
 CC PROSITE: PS00661; BAND_4.1_2; FALSE_NEG.
 CC PROSITE: PS50057; BAND_4.1; 1.
 CC PROSITE: PS50096; IQ; 2.
 CC PROSITE: PS50093; PH_DOMAIN; 2.
 CC KW Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat.
 CC FT DOMAIN 1 727 HEAD OR MOTOR DOMAIN.
 CC FT 742 763 IQ 1.
 CC FT DOMAIN 764 787 IQ 2.
 CC FT 788 817 IQ 3.
 CC FT DOMAIN 1206 1304 PH 1.
 CC FT 1386 1491 PH 2.
 CC FT DOMAIN 1691 1975 BAND 4.1-LIKE.
 CC FT 800 941 COILED COIL (POTENTIAL).
 CC SQ SEQUENCE 2052 AA; 235837 MW; 43DF13424B4BD28 CRC64;
 Query Match 92.8%; Score 9961; DB 1; Length 2052;
 Best Local Similarity 92.7%; Pred. No. 0;
 Matches 1909; Conservative 76; Mismatches 67; Indels 12; Gaps 6;
 QY 1 FCLQSTRVWLENQGHFPSTVNSCAEGIVFRIDYGVFTYKQSTITTHQKVTAMPTNEE 60
 DB 4 FPEESTRWLENQGHFPSTVNSCAEGIVVQTDVGVFTYKQSTITTHQKVTAMPTNEE 63
 QY 61 GVDMASTLEHAGSINTNLFQRYKRNQIWTYIGSILASVNPYCFIAGYEPATMEQYSR 120
 DB 64 GVDMAATLEHAGHIMNLFQRYKRNQIWTYIGSILASVNPYCFIAGYEPATMEQYSR 123
 QY 121 RHIGELPHFAINAECCRCAMKRDHNCILIKGSGAGKSTETYLILKPLFVSISQOSLE 180
 DB 124 CHLGEPLPHFAINAECCRCAMKRDHNCVILSGSGAGKSTETYLILKPLFVSISQOSVD 183
 QY 181 LSLKEKTSCEVRALIESPIMEAFGNATVNNNSRFGKVFQJNICQKGINQGRIVDC 240
 DB 184 LSSKEKTSCEVRALIESPIMEAFGNATVNNNSRFGKVFQJNICQKGINQGRIVDC 243
 QY 241 ILSSGNRVRRNGPGRNTHITVALLAGIEHREBEFYISTEENHYLNQSGCVERKXTSD 300
 DB 244 LL-EKRVRRNGPGRNTHITVALLAGIEHREBEFYISTEENHYLNQSGCVERKXTSD 302

Db 821 QACRAVLVKAARHRLMAVITVOAAYARGLIARLRHRLRVEYORRLAEERMLAESEKL 880
 QY 830 KRE-----EEREREREREAELRAQOEETRKQOELALOKOKEAELTRELEKOK-- 881
 Db 881 RKMMAKKAKEERERKQERLADAREDAE-----RELKEKEARKEKELLEOMEKARHE 935
 QY 882 ---ENKQVEIL-----RLEKEIEDLORMEQOELSLEASLOKLQERDO 924
 Db 936 PINHSMVMDKMPFPLTSGSLPGQEQAPGCFEDLERGR----- 974
 QY 925 ELARLEECRAQOEFLESINPEIDECPNIRSLSGSEFSESLAESACEKKNPNTS 984
 Db 975 ---REWEEDVDALPLP---DEDEEDSEYFAFAATYFOGTTTHS-YTRRP--LK 1023
 QY 985 QP---YEEBEYDE-----GFEADDDAFKSPNPSHSGSDORTSGIRTSDDS 1028
 Db 1024 QPLHYHDDBGDLAALAVITLIRFMG-----DLPEPKH-----TMSDQ 1064
 QY 1029 SEEDPYM-----NDIVPTSPSADST-VLLAPSVODSGSLHN 1064
 Db 1065 SEKI PVMTKIYETLGKTYKRELQALQGEGETQLPEGQKTSYRHKLVHLTLKKKSRLTE 1124
 QY 1065 S-----SSGSESTYCMQNGDLPSPD-----GDYDYOQDDVEDDAIS 1102
 Db 1125 EYTKRLNDGESTYQGNMSMEDRPTSNLEKHLIIGNGILRALRDLITQISKQ---LTH 1181
 QY 1103 GSSVTFNSYSGSQSPDYRCSTVGTYNSS-----GAYRFSEGAQSGSPED 1146
 Db 1182 NPS---KSSYARGMILVSLC-VGCFAPSEKFKYKLNFIHGPRGAPACEBELRRTTFVN 1237
 QY 1147 SEEDPFRFPTDELSTYRDSVYSCTLPYFHSFLYMKGGMLN-----SWKRWG--- 1196
 Db 1238 G-----TRTOPPSMELQATSKSKPIMLP---VTMEDGTTKTLTDSATTABELCNALA 1288
 QY 1197 ---VLDD---EFLFMRSKQELKOGMLH-----KKGSSSTLSRMKKRM 1237
 Db 1289 DKISLDRGFSLYIALFDVSSLSGSDHMDAISQCEQYAEQQA---QERNAPMWL 1344
 QY 1238 FVLROKLMFYENSEEKLKGTV-----EVRTAKE--IIDNTTKENGIDI-- 1280
 Db 1345 F-FRKVFPTPMHNPSEDNVAITNLIYQGVYGVAFGEYRCKEDDLAELASQYFVYGSSE 1403
 QY 1281 IMADRTPHLIAE-----SP-EDASQWFSVLSQVH---ASTDEIQEMHEQANPON 1327
 Db 1404 MLIERLISLVPTIPIPREITPLKMLEKMQALAIKAKGIYARRTDSQVKEVDVNVYAR 1463
 QY 1328 AVGTLDVGLIDSCASDSPRPSFIIITANRVLHCNADPREMHMITLLOKSKDITV 1387
 Db 1464 FKMPFLFSRFEYAYKFSGPPLPKSDVIAVANMTGVYFVDEQBOV---LLELSPEEIMA 1518
 QY 1388 ---EGQEFIV-----RGWLHKEVKNSPKMS--SLKLKRMFVLTHNSLD 1426
 Db 1519 VSSSRRCRVLISLGCSDLCATQOSRAGLTGAPGSPSCSGCTGMAMPFLATIKD 1578
 QY 1427 YKSSSEKALAKTLVNSLCSVPPDEKTFKCTGW---NVTYGRKHCYRLYTK-- 1479
 Db 1579 EYFTFSNMEDRIDLVVTFP---EGLRKRSKTVVVALQDNPAGSESGSLFPAKD 1631
 QY 1480 ---LLEAT-----RMSVIONVTDTPKPIDT-----PQOQL----- 1509
 Db 1632 LIILDDHDTGEQVNSGANGINERTKORGDFPTDCVYVMTPTLTPREIALTVMTPDOR 1691
 QY 1510 ODI-----KENCLNSDVVEOIIYKRNPIRYTHNP 1538
 Db 1692 QDVVRLQLQRTAPEVRAKPYTLEFSYDYFRPPPHITLSRVVNSAKRGDRLMSTTRP 1751
 QY 1539 LHSPLPLPYGDINLMLDKGYTTLODAIKIFNSL-----QOLESMDPIPIIO 1589
 Db 1752 LKQALLK-----KILGSEESQEAQMAFVAVLKMGDYPSKRMASVNE---LMD 1797
 QY 1590 GILQTHDLRLADELYCOLIQ--TNKVPHPSGVNLSWQVLJTLGSLTFPSRGLIKY 1647
 Db 1798 QIFEMALKAPLKDDEAVOILKQLTDNHRIYSEERG---WELLMLCTGLFPSPNLLPH 1853

QY 1648 LKFHLKRIEOPPGTEMERYALFTYESLKK---TKCREVPSRDEIALIHR-QEMTSTV 1703
 Db 1854 VORFLQGS-RKICP-----LAIDCLQRLQKALRNSRKYPHVLVEVAIQHTTQIDPHKV 1906
 QY 1704 YCHGGSCKITITNSITTTAGEVVEKILRGLAMEDSRNML-----FEY 1746
 Db 1907 YPPDDTDEAFEVESSTKADPCONTASRLLLKSSBG-FSLPVKIXADKVISVENDFFPDF 1965
 QY 1747 NGHVDKAIERSRVVADVLAKEFLKLAATSEVGDLPKPFYF--KLUYCELDTPNP-KDSV-E 1802
 Db 1966 VHLHLDWIKKAPRIDGI-----VPSLITYQVFFMKLM---TTVPGADPMAD 2010
 QY 1803 FAFMEQAEAVIHGHHPAPEENLOVAL--RLQYLOGDYTLHAALPBLEEVSLORLK 1860
 Db 2011 SIFHYQELPKYLRGVHKCTREEVLOGLAIYRVKF-EEDKSYFPSIPXL----- 2059
 QY 1861 ARISOSTKFTTCELEKRRITSFLEGLRBSRPTSSVVRQKVEEQMLDMWKEVSSAR 1920
 Db 2060 -----LR-----ELPQDLIRQVSPDW-----K 2078
 QY 1921 ASIIDKMRKFOGANOBOAMAKYMALIKEMPGYSTLPDY-EGKEGFPOLMVGSAV 1979
 Db 2079 RSIIVAFNKHAKSKSEAKLAFKLIKMPFTGSAFVEYKQTTBENFEPILIAINKGV 2138
 QY 1980 SVYKGEGRPLEVFQYEHILSPGAPLANTYKIV-----VDERELLFETSEVVDAKLMKA 2034
 Db 2139 SLIDRPTDILTTHFTTKISNMS--GNTYPIITIGNLVRSGLKLCETSLGKMDLLTS 2196
 QY 2035 YISMVYKRYSTTRASNSGSSR 2057
 Db 2197 YISQWL-----TAMSKQRNSR 2212
 RESULT 4
 MYTA HUMAN
 ID MYTA HUMAN STANDARD; PRT, 2215 AA.
 AC Q13402; Q92821; Q92822; Q13321; P78427; Q14785;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin VIIa.
 GN MYO7A OR USH1B.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 CX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1; 5; 6 AND 7), AND DEVELOPMENTAL
 RP EXPRESSION.
 RC TISSUE=Retina;
 RX MEDLINE=96194949; PubMed=8622919;
 RA Weil D., Levy G., Sahly I., Levi-Acobas F., Blanchard S.,
 RA El-Amraoui A., Crozet F., Philippe H., Abitbol M., Petit C.,
 RT "Human myosin VIIa responsible for the usher 1B syndrome: a predicted
 RT membrane-associated motor protein expressed in developing sensory
 RT epithelia".
 RL Proc. Natl. Acad. Sci. U.S.A. 93:3232-3237(1996).
 RP [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 2, 3 AND 4).
 RC TISSUE=Testis;
 RX MEDLINE=97038686; PubMed=884267;
 RA Chen Z.-Y., Hasson T., Kelley P.M., Schwender B.J., Schwartz M.F.,
 RA Ramakrishnan M., Kimberling W.J., Mooseker M.S., Corey D.P.;
 RT "Molecular cloning and domain structure of human myosin-VIIa, the gene
 RT product defective in usher syndrome 1B".
 RL Genomics 36:440-448(1996).
 RN [3]
 RP SEQUENCE OF 166-196 FROM N.A.
 RC TISSUE=Epithelium, Leukocyte, and Liver;
 RX MEDLINE=94294418; PubMed=8022818;
 RA Bement W.M., Hasson T., Wirth U.A., Cheney R.E., Mooseker M.S.;
 RT "Identification and overlapping expression of multiple unconventional

RT myosin genes in vertebrate cell types.";
 RN Proc. Natl. Acad. Sci. U.S.A. 91:6549-6553(1994).
 [4]
 RC SEQUENCE OF 1-1075 FROM N.A.
 RP TISSUE=Testis;
 RX MEDLINE=96003870; PubMed=7568224;
 RA Hassan T., Heintzelman M.B., Santos-Sacchi J., Corey D.P.,
 RT Mooseker M.S.;
 RT "Expression in cochlea and retina of myosin VIIa, the gene product
 RT defective in Usher syndrome type 1B.";
 RL Proc. Natl. Acad. Sci. U.S.A. 92:9815-9819(1995).
 [5]
 RP SEQUENCE OF 96-564 FROM N.A. (ISOFORM 1), AND VARIANTS USH1B.
 RC TISSUE=Retina;
 RX MEDLINE=95174880; PubMed=7870171;
 RA Weil D., Blanchard S., Kaplan J., Guilford P., Gibson F., Walsh J.,
 RT Mburu P., Varela A., LeVilliers J., Weston M.D., Kelley P.M.,
 RA Kimberling W.J., Wagenaar M., Levi-Acobas F., Larget-Piet D.,
 RT Munnich A., Steel K.P., Brown S.D.M., Petit C.;
 RL "Defective myosin VIIA gene responsible for Usher syndrome type 1B.";
 RN Nature 374:60-61(1995).
 [6]
 RP SEQUENCE OF 79-578 FROM N.A.
 RX MEDLINE=97224487; PubMed=9070921;
 RA Kelley P.M., Weston M.D., Chen Z.-Y., Orten D.J., Hassan T.,
 RA Overbeck L.D., Plint J., Talmadge C.B., Ing P., Mooseker M.S.,
 RA Corey D.P., Sumegi J., Kimberling W.J.;
 RT "The genomic structure of the gene defective in Usher syndrome type 1b
 RT (MYO7A).";
 RL Genomics 40:73-79(1997).
 [7]
 RP DEVELOPMENTAL EXPRESSION.
 RX MEDLINE=96440426; PubMed=8842737;
 RA El-Amraoui A., Sallly I., Picard S., Sahel J., Altbol M., Petit C.,
 RT "Human Usher 1B/mouse shaker-1: the retinal phenotype discrepancy
 RT explained by the presence/absence of myosin VIIA in the photoreceptor
 RT cells.";
 RL Hum. Mol. Genet. 5:1171-1178(1996).
 [8]
 RP VARIANTS USH1B H-212; C-212; H-302; Q-468 INS; L-503 AND Q-450.
 RX MEDLINE=97055890; PubMed=8900236;
 RA Weston M.D., Kelley P.M., Overbeck L.D., Wagenaar M., Orten D.J.,
 RA Hassan T., Chen Z.-Y., Corey D.P., Mooseker M.S., Sumegi J.,
 RA Cremer C., Mueller C., Jacobson S.G., Gorin M.E., Kimberling W.J.;
 RT "Myosin VIIA mutation screening in 189 Usher syndrome type 1
 RT patients.";
 RL Am. J. Hum. Genet. 59:1074-1083(1996).
 [9]
 RP VARIANTS USH1B ARG-214; ASP-397 AND THR-826, AND POLYMORPHISMS.
 RX MEDLINE=98018437; PubMed=9382091;
 RA Adato A., Weil D., Kalinski H., Pel-Or Y., Ayadi H., Petit C.,
 RA Korostoffsky M., Bonne-Tamir B.;
 RT "Mutation profile of all 49 exons of the human myosin VIIA gene, and
 RT haplotype analysis, in Usher 1B families from diverse origins.";
 RL Am. J. Hum. Genet. 61:813-821(1997).
 [10]
 RP VARIANTS USH1B ARG-25; SER-955 AND GLU-2137, AND POLYMORPHISMS.
 RX MEDLINE=97156218; PubMed=9002678;
 RA Levy G., Levi-Acobas F., Blanchard S., Gerber S., Larget-Piet D.,
 RA Chapal V., Liu X.-Z., Newton V., Steel K.P., Brown S.D.M., Munnich A.,
 RA Kaplan J., Petit C., Weil D.;
 RT "Myosin VIIA gene heterogeneity of the mutations responsible for
 RT Usher syndrome type 1B.";
 RL Hum. Mol. Genet. 6:111-116(1997).
 [11]
 RP VARIANT DFNB2 PRO-244.
 RX MEDLINE=97315827; PubMed=9171833;
 RA Liu X.-Z., Walsh J., Mburu P., Kendrick-Jones J., Cope M.J.,
 RA Steel K.P., Brown S.D.M.;
 RT "Mutations in the myosin VIIA gene cause non-syndromic recessive
 RT deafness.";
 RL Nat. Genet. 16:188-190(1997).
 [12]

RP VARIANT DFNB2 ILB-599.
 RX MEDLINE=97315828; PubMed=9171833;
 RA Weil D., Kuesel P., Blanchard S., Levy G., Levi-Acobas F., Drita M.,
 RA Ayadi H., Petit C.;
 RT "The autosomal recessive isolated deafness, DFNB2, and the Usher 1B
 RT syndrome are allelic defects of the myosin-VIIA gene.";
 RL Nat. Genet. 16:191-193(1997).
 [13]
 RP VARIANT DFNB1 886-ALA-LYS-888 DEL.
 RX MEDLINE=98016404; PubMed=9354784;
 RA Liu X.-Z., Walsh J., Tamagawa Y., Kitamura K., Nishizawa M.,
 RA Steel K.P., Brown S.D.M.;
 RT "Autosomal dominant non-syndromic deafness caused by a mutation in the
 RT myosin VIIA gene.";
 RL Nat. Genet. 17:268-269(1997).
 [14]
 RP VARIANTS USH1B PRO-651 AND GLN-1602.
 RX MEDLINE=98386090; PubMed=9718356;
 RA Liu X.-Z., Hope C., Walsh J., Newton V., Ke X.M., Liang C.Y., Xu L.R.,
 RA Zhou J.M., Trump D., Steel K.P., Bunday S., Brown S.D.M.;
 RT "Mutations in the myosin VIIA gene cause a wide phenotypic spectrum,
 RT including atypical Usher syndrome.";
 RL Am. J. Hum. Genet. 63:909-912(1998).
 [15]
 RP VARIANT USH1B PRO-1087.
 RX MEDLINE=99294608; PubMed=10364543;
 RA Adato A., Kalinski H., Weil D., Chatb H., Korostoffsky M.,
 RA Bonne-Tamir B.;
 RT "Possible interaction between USH1B and USH3 gene products as implied
 RT by apparent digenic deafness inheritance.";
 RL Am. J. Hum. Genet. 65:261-265(1999).
 [16]
 RP VARIANTS USH1B AND POLYMORPHISMS.
 RX MEDLINE=99140257; PubMed=10094549;
 RA Janacke A.R., Meins M., Sadeghi M., Grundmann K., Apfelstedt-Sylla E.,
 RA Zrenner E., Rosenberg T., Gal A.;
 RT "Twelve novel myosin VIIA mutations in 34 patients with Usher syndrome
 RT type I: confirmation of genetic heterogeneity.";
 RL Hum. Mutat. 13:133-140(1999).
 [17]
 RP VARIANTS USH1B LYS-1170 AND CYS-1719.
 RX MEDLINE=99374392; PubMed=10447383;
 RA Cuevas J.M., Espino C., Millan J.M., Sanchez F., Trujillo M.J.,
 RA Ayuso C., Benito M., Najera C.;
 RT "Identification of three novel mutations in the MYO7A gene.";
 RL Hum. Mutat. 14:181-181(1999).
 [18]
 CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
 CC THEIR HIGHLY DIVERGENT TAILS ARE PRESUMED TO BIND TO MEMBRANOUS
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS. IN
 CC RETINA, MYOSIN VIIA MIGHT PLAY A ROLE IN TRAFFICKING OF RIBBON-
 CC SYNAPTIC VESICLE COMPLEXES AND RENEWAL OF THE OUTER PHOTORECEPTORS
 CC DISKS. IN INNER EAR, IT MIGHT MAINTAIN THE RIGIDITY OF STEREOCILIA
 CC DURING THE DYNAMIC MOVEMENTS OF THE BUNDLE. INVOLVED IN HAIR-CELL
 CC VESICLE TRAFFICKING OF AMINOGLYCOSIDES, WHICH ARE KNOWN TO INDUCE
 CC OTOTOXICITY (BY SIMILARITY).
 CC -1- SUBUNIT: MIGHT HOMODIMERIZE IN A TWO HEADED MOLECULE THROUGH THE
 CC FORMATION OF A COILED-COIL ROD.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE). IN THE PHOTORECEPTOR
 CC CELLS, MAINLY LOCALIZED IN THE INNER AND BASE OF OUTER SEGMENTS AS
 CC WELL AS IN THE SYNAPTIC ENDING REGION.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 7 ISOFORMS; 1 (SHOWN HERE), 2, 3,
 CC 4, 5, 6 AND 7; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE PIGMENT EPITHELIUM AND THE
 CC PHOTORECEPTOR CELLS OF THE RETINA. ALSO FOUND IN KIDNEY, LIVER,
 CC TESTIS, COCHLEA, LYMPHOCYTES. NOT EXPRESSED IN BRAIN.
 CC -1- DEVELOPMENTAL STAGE: DETECTED IN OPTIC CUP IN 5.5 WEEKS-OLD
 CC EMBRYOS. EXPRESSED IN RETINAL PIGMENT EPITHELIUM, COCHLEAR AND
 CC VESTIBULAR NEUROEPITHELIUM, AND OLFACTORY EPITHELIUM AT 8 WEEKS. AT
 CC 19 WEEKS, PRESENT IN BOTH PIGMENT EPITHELIUM AND PHOTORECEPTOR
 CC CELLS. AT 24-28 WEEKS, EXPRESSION IN PIGMENT EPITHELIUM AND
 CC PHOTORECEPTOR CELLS INCREASES. PRESENT IN PIGMENT EPITHELIUM AND
 CC PHOTORECEPTOR CELLS IN ADULT.

CC -1- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF USHER SYNDROME TYPE 1B
 CC (USH1B), ALSO KNOWN AS USHER SYNDROME NON-ACADIAN VARIETY. USHER
 CC SYNDROME TYPE 1 IS AN AUTOSOMAL RECESSIVE DISEASE CHARACTERIZED BY
 CC PROFOUND CONGENITAL SENSORINEURAL DEAFNESS, CONSTANT VESTIBULAR
 CC DYSFUNCTION AND PREPUBERTAL ONSET OF PROGRESSIVE RETINITIS
 CC PIGMENTOSA LEADING TO BLINDNESS. USHER SYNDROME IS THE MOST COMMON
 CC CAUSE OF COMBINED DEAFNESS AND BLINDNESS IN DEVELOPED COUNTRIES.
 CC USH1B APPEARS AS A PRIMARY CYTOSKELETAL PROTEIN DEFECT.
 CC -1- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL RECESSIVE
 CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNB2 OR NSRD2) WITH
 CC VARIABLE VESTIBULAR DYSFUNCTION AND VARIABLE AGE OF ONSET.
 CC -1- DISEASE: DEFECTS IN MYO7A ARE THE CAUSE OF AN AUTOSOMAL DOMINANT
 CC FORM OF NONSYNDROMIC SENSORINEURAL DEAFNESS (DFNA11) WITH ONSET
 CC AFTER COMPLETE SPEECH ACQUISITION AND SUBSEQUENT GRADUAL
 CC PROGRESSION.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 IQ DOMAINS.

Query Match 17.1%; Score 1831; DB 1; Length 2215;
 Best Local Similarity 26.1%; Pred. No. 2e-76;
 Matches 632; Conservative 379; Mismatches 822; Indels 592; Gaps 87;

QY 4 QGNRWVL-RENGHPESTVSCAEGIVFRTDGVFTYKQSTIT- - - - -QKTAH 55
 DB 6 QGVHVMDELRLGDFVPI- - - - -GAVVKLDSGQVQVDEDENEHMSFQNAATHIKPMH 60
 QY 56 PTNEGVDMASLTSLHGSIMYNLQRYKRNQIWTYIGSILASVNPYOPIALYEPATM 115
 DB 61 PTVHGVEDMIRLGDINEAGILNLRDHLIYTGSLIVAVPYQ-LLSISPEHI 119
 QY 116 EYSRRLGELPHIFAIANECTRLMKRHDNOCILKESGAGKTESTYLILKFLSVIS 175
 DB 120 RQYTNKKTIGMPHIFAIADNCFNMKRNRSRDOCCIISGSGAKTESTYLILQFLAAS 179
 QY 176 QOSLESLAKETSCVERALLESSEPIEAFGNATVNNNSRGRKFPQNLICQKNIOG 235
 DB 180 GQH- - - - -SWIEQVLEATPILEAFGNATKIRNDSSRFKXIDHFNRGALIEG 230
 QY 236 RYDCILSSQNRVVRQNGERNYHIFAYALAGLEHEBEREFYLTSENYHYLIQSGVED 295
 DB 231 KIQGYLL-EKSRVCRQALDERNHVFCMLGMSBQKKLGLGASDVNYLYAMGCITC 289
 QY 296 KITSDOSEPREVTAMDVQFSKEVREVRLLAGILHGNIF- - - - -FTACGAOVSEK 350
 DB 290 EGRVDSOEYANIRSAKVMLEFDTENWEISKLAALILHGNLQYEARTFENLDACEVLS 349
 QY 351 TATGRSAILGLPTQTLDTQSRMFLRGEELPLPNQOAVDSDSLAMALYACCFEN 410
 DB 350 PSLATASILEVNPDPMLSCLTSTRLITRGETVSTPLSRQALDVRDAFYKIGRLPYW 409
 QY 411 VIKINSRI--KGNEDFK- - - - -SIGILDFGFENFEVNHFEQFNINYANEKLOEYFNKH 464
 DB 410 IYDKINAIYKPPSOVKNSRSRISGLIDIFGFENFVANSFEQICINPANEHLQGFVRHV 469
 QY 465 FSLSELEYSREGVLEDIDINGECULDI-EKKGLALINEESHFPQATDSTLLEKHA 523
 DB 470 FKLEOEYDLESIDWHLIEFTDQDALDMANKPMNIIISLIDESKFPKTDITTMHLKHN 529
 QY 524 SOHANNHFYKPRVAVNN- - - - -FGVRYAGEVOYDVGRILEKXRPDRDILNLLESPR 579
 DB 530 SQKLANIYTPK- - - - -NNHETQGINHFAQIVYEETGFLERKRDLDHGDIDLVSSRN 586
 QY 580 DFIYDLFEHVSRRNODTLKCSKRR- - - - -PTVSSQFQVDSLHSLMATTSSNPFVRCIK 637
 DB 587 KPIKQIF- - - - -QADVAMGAEIRKSPITSSQF-RSLBELMRTYLAGCQFFVRIC 637
 QY 638 PNNQKRPDQAVLNLQRLYSGLMLETVRIRKAGYAVRPFQDFYRYKYKMLNLA- - -L 694
 DB 638 PNFKKKPMLEFDRHLCYRQLRYSGMETIRIRRAGYDIRYSFVEFVRYVLLRGLGVPAVK 697
 QY 695 PEVVRGKCTSLDLAYSANSEMOLGKTVLRESLOKLEKRESEVSHAAMVIRAHVIG 754
 DB 698 QGDLRGTCORMAEAVLGTDHWDQIGKTKIFLKDHHMLLEVERDKAITDRVILLQKVRIG 757

QY 755 FLARQYKRVLYCVIIQ- - - - -KXY- - - - - 775
 DB 758 FDRSNFLKKAATLIGNHGRCHNCRKYGMRLGFLRLQALHSRKLHQYRLARQRI 817
 QY 776 - - - - -RAFLRRFLHLKKAAYFQKQRLQIARRVYRLAER- - -REGEKKQKE 826
 DB 818 IOFQARCAVYLRKAFRRHMLVLTVOAYARIMARRLHORIRAYVLRLEKEKRLAE 877
 QY 827 EKKKE- - - - -EEREREREREAELRAQEEETRKQOELALQKQEAELTRELEKO 880
 DB 878 ELKREMSAKKAEBAERKHQRLQAREDA- - - - -RELKEKAARKKELLQOMERA 932
 QY 881 K- - - - -ENKQVEIL- - - - -RLEKEIEDIORMKEQOELSTEASLOKQER 921
 DB 933 RHEPVNSDMVQKMGFLGTSGGLPGQEQAPSGFEDLERGRER- - - - -VEED 981
 QY 922 RQELRLEEEACRAQOEFLESINDELDECVRNIRLSGSGSEFSSELAESACEKPNF 981
 DB 982 LPAALPLPED- - - - -EEDLSEYKPAKF- - - - -AATYQGTTHSYTRRP- - 1021
 QY 982 NFSQ- - - - -YBEEVDE- - - - -GFEADDAFQDSPNPSHGHSDQRTSGIRTS 1025
 DB 1022 -LKQPLHYDDGDLAALAVITTLRNG- - - - -DLPEPKYH- - - - -TAM 1061
 QY 1026 DDSSEDPYMDTVVPTSPSADSTVLA- - - - -PSVQDSGLHNS- - - - - 1065
 DB 1062 SDGSEKIPVMTKIYETLGKTKYRELQALQGEBAQLPEGQKSSVRRHLVHLTLKKSK 1121
 QY 1066 - - - - -SGSESTYCPQANAGLPSPD- - - - -GQDYQODDYEDA 1099
 DB 1122 LTEEYTKRLHDESTVOGNSMLEDRPTSNLEKHLPIIGNGILRPALEIYQISQ- - 1178
 QY 1100 ITSGSVTFPSNGSGQSPDYRCVSGTYNS- - - - -GAYRPSSEGAQSS 1143
 DB 1179 LTHNS- - - - -KSYAAGMILVSLC-VGCRAPSKFPYKILRNFHGGPGRGAPRCEBRLKRT 1234
 QY 1235 FVNG- - - - -TRQPSWLELQATKSKKPIMLP- - - - -VTFMDGTYLLTDSATYAKELCN 1285
 DB 1190 SKWRKCVLKD- - - - -ETPLMFPSKQELKQGLH- - - - -KKGGSSTLSRW 1233
 QY 1286 ALADK-ISKORFGSLYIALFDKXVSLGSGSDHWDMAISOEOYAKQGA- - -QERNA 1340
 DB 1234 KRWFLRQSKLMYENDSEELKGTV- - - - -EVRTAKE- - -IIDWTKEGND 1279
 QY 1341 PWRLF-FRKEVTPMHSSEEDVATNLIYQVVRGVKGEYRCEKEDDLAELASQYFVD 1399
 DB 1280 I- - - - -IMADRTFLHIAESPED- - - - -ASQVSVLSQVH- - - - -ASTDOELQEMHDEQA 1323
 QY 1400 YSEMIERLNLVLPYIIPREITPLKTLKKAQALAIAAHKKGIYAQRRTDAQYKVEDV 1459
 DB 1324 NQONAVGLTDLVLIISVCASDSPDRPNSFVITITARVHLHCNADTDEEHNMHTLLQRSKG 1383
 QY 1460 SYARFKWPLFSRFYBAKFGSPSLPKQDVIVAVMGTGYFVDEBOV- - - - -LLELSP 1514
 DB 1384 DTRVEGOEFIVRGMLHKEVKN- - - - -SPKMS- - - - -SLKLKRWFLVTH 1422
 QY 1515 EIMAVSSSEKCVWLSTLGSDDLGAAPHSAGMLTPAGBCSFCWSCRAKTAAPFTLAT 1574
 DB 1575 IKGDEYTFSSVAEDRIDLVVTF- - - - -EGLKRSKYVALADQNPAPAEESGFLSF 1627
 QY 1478 TK- - - - -LNEAL- - - - -RMSVYQNTVDTAAPIDT- - - - -PTQOLI- - - - - 1509
 DB 1628 AAGDIIIDHDYGEVNSGANGINERTKORGDEPTQCVVMPTVTPPREIVALVMT 1687
 QY 1510 - - - - -ODI- - - - -KENCINSDVVEQIKRNPILRY 1534
 DB 1688 PQRQDVVRLDLQRTABEVRAKPYTLLEFSYDVRPPPKHLSKVMVMSKAGKQRLMSH 1747

FT	DOMAIN	2848	2934	SH3
FT	NP_BIND	1289	1306	ATP (POTENTIAL).
FT	VARIANT	1779	1779	C -> Y (IN SH2).
FT	CONFLICT	1330	1331	MISSING (IN REF. 2).
FT	CONFLICT	1579	1579	L -> R (IN REF. 3).
FT	CONFLICT	1955	1972	MISSING (IN REF. 2).
FT	CONFLICT	2077	2077	L -> M (IN REF. 2).
FT	CONFLICT	2139	2139	L -> P (IN REF. 2).
FT	CONFLICT	2953	2953	V -> A (IN REF. 2).
SO	SEQUENCE	3511 AA;	395533 MW;	38C962P98A2D395B CRC64;
Query Match 14.8%; Score 1593; DB 1; Length 3511;				
Best Local Similarity 24.8%; Pred. No. 3; Le 65;				
Matches 609; Conservative 373; Mismatch 836; Indels 642; Gaps 81;				
QY	59	EGVDNMASTLHSGSIMYNL	FORKRNQIWTYIGSLASVNPQYPLAGYEPATMBOY	118
DB	1205	EDGVEDMTQDLBQJOTYLANL	KTRFERNLITYTIGSLVSNPRMR - IYGPQVOY	1265
QY	119	SRRLGELPHI	PAIANECYRCLMKRHDNOCILKGESGAGTESTKILKFLSVISQOS	178
DB	1264	SGRALGEPNHL	PAIANLAFKMLDAKONOCYITISGSGSGTETATKILRCLAMQR	13222
QY	179	LELSIKETSCVERAL	LESSPIMEAFGAKTYVNNNSRFGKFGVQNLNCOGNGIOGRIV	238
DB	1323	-----ADVMOQIK	LEATPLLEFGAKTYVNDNSSRFGKFGVFIPL - REGVICAT	13737
QY	239	DCILSSQNRVVR	NGERNYHIFYLLAGLHEHEEREFTLSTPEVNYHLNOSGCEVDTI	298
DB	1374	SOYLEKSRIV	FOAGNERNYHIFYLLAGLPAQLRQASLSQEAFTYYLLNOSGCEINAK	14333
QY	299	SDQSFREVITAM	DVMOFSKEEVEVSLLGLILGNIEF ----TAGGAQVSFKTAL	353
DB	1434	SDADDFRILAMEV	LGTSEDOOSIFRILASILHLGNVFEKHEPTDAQEVAVSAREI	14939
QY	354	GRSEILGLDPTQ	LDTALTOFSMFLRGEILITPLNVOQAVDSRSLAMALVACSEWIK	413
DB	1494	QAVELLQVSP	EQKATITFKVETIRKITPLPLVESAVARDIAVLVALFEGWILT	15535
QY	414	KINSRIKQNE	DPKSGIGLIDIFGFENFEVNHFEQFINVANEKLOEYFNKHI	FSLEOLEYS 473
DB	1554	RVMALVSKQD	TLTAIIDYGFEDLSNSPQCLINAVENLQYLPKXIYFOEQDEYI	16133
QY	474	REGLVEMEDIM	INDGECIDLE -KKLGLALINEESHFPOATDSTLLEKLSOHNANHY	532
DB	1614	REOMDMREI	APADNPICNLISLKRYGILRTIDQCCPQATDHTFLQKCHYHNGANDLY	16737
QY	533	VKPRVAVN	FGVKYHAGVOYDVAGILKRNKDPFDLDLNLBRESRPFYIDLPENHVS	592
DB	1674	SKPKPRLEP	EFITKHVAGKTYOVNHFKLDKNDQVQDVLDFVHSRTVVNHLFESSHAQ	17333
QY	593	NNODTLKCGSK	----HRRPTVSGQKVDLSLMAATSSNPFPRCIRKPMQKRPDOFD	648
DB	1734	TAPRLGLGSS	SITRLRYKATVAAKQ -OSLIDLVEKMRCHNPLFRCLKPNHKEBGLFE	17992
QY	649	QAVVLNQLRSG	MLTVIRKAGAVRRPFQDFYKRYKLMR -MLALBEDVKGKSTLL-	706
DB	1793	PDVMAQAQ	LRYSGLVETVIRIRKGGFVRLPFQVIFDRYCVLAKLVNAD -GDMCVSLIS	18515
QY	707	QLYDASNE	MOGLKTVLRLSLEQKLEKREBEVSHAAWYIRAHVGLFRLARKOYRKVLY	766
DB	1852	RLCTVTPMY	VVIGSKFLKEHNLQLESMEERVONRAALTLORLARGFLOMRHSRLR	19111
QY	767	CVVILIQNKY	RAFLRRRFLHKKAIIVEOKDRLQGLIARVYROLLAEKREO -EEKKQOE	825
DB	1912	KIILQSGAR	GLFARQYOMKQOSLTKRSLVHTYVNNRRYKLABORRRQAOEWLMEQ	19717
QY	826	EEKKKREBE	REEREREAELRAQOEETKQOELMALQOSQKEAEVLT --RELKQKE	882
DB	1972	ELSKRELV	-----RVNHLVRA -----EVALLQAAAGLKLSSGPRVAVUNAP	2015
QY	883	NKQVEILIR	LEKEJEDLOKMEQOELSTEASLOKLERRDOELRLE -EERCAAOEFL	941

Db	2016	RLQAEPCVTLRPLDINNTP-MAKIFRCHFKERPSFQMLTVPLKMLPTRLRPHVHNHAIVSK	2074
Qy	942	ESLNFDEIDECVNRNIEKSLGSGSEFSSLEASACEK-----PN-ENFSOPY-	987
Db	2075	LILRFMG-DPHLHOTQEMILGNYIVHQGLVEPALRDEILAQLANQVWRNPAAVYSKSGWL	2133
Qy	988	-----PEEVEDE-----GPEADDDAKPSPNSEHGHSQKRSIGRTS	1025
Db	2134	LAAACISGFAPSPHLDFKLLKPVSDYQGNQFOA-----VQHRLLQAMGSGAART	2183
Qy	1026	DDSEEDPWNMDTVPTSPSPADSTVLLAPSVQDSGLHNHSSGESTYCMENQAGDLESPD	1085
Db	2184	FPFPQO-----LEMTAIOEKAMALDVSCFNGDQBSCPHTNISTGSAV-----AGDILKHR	2233
Qy	1086	GDYDODDDYEDGAITSGSVTSNSYGSQWSP-----DYRCSV-----GTYN	1128
Db	2234	G-----LADG--WRGWTIVAMGN--GVQWAEIAGHDYLDVLDSLELLRDFPROKSYF	2281
Qy	1129	SSGAFRRSSGAGSFFDSEEDPSPRDTDELSYRDSYSCTLYEHSFLYMK--GG	1186
Db	2282	IYGA-----EGPLAGRGDTRGVFGNCWDSBDPTPRPOPDHVAKMPELDGYSCHKDGT	2336
Qy	1187	LNMSMKRRW-----CYLKDETFLMFRSQO	1210
Db	2337	NGETBAOKRMSNROAVDSIGESTYPPRELDGYLDSLFDPLVACGDADLEKPTAIAYMK-	2355
Qy	1211	EALKQGMIAHKGGGSSST--LSRRNMK-----KRWFLRSQKLMYEEND	1251
Db	2396	-----GGGQGGGGGGSTSEDTSRRPAPKPIRGDLASTLALQGAFTHQAVLAREMT	2450
Qy	1252	-----SEELKSGT-----VAVR-----TAKEITDNTTKEN-----	1276
Db	2451	LQALALQOQPLASAIRPOLBERPLAPARPKTVNGTPRPAKPVLPVLRTPQSWABGSAKA	2510
Qy	1277	-----GHDIWADRTF-HLIAESPEDASQWFSVLSQVH-ASTDQEIOM-----	1318
Db	2511	PKIJSKVAVPIILAQDWTAPESISASPELYR--YSTLNSEHPFOPTQOINSITIQYQOPR	2568
Qy	1319	-----HDE-----QANPONAVGTLDVGLDSV-----	1340
Db	2569	WAGHPEARRTDGGKVFRRPRDPHHEALMILKGQTOGLAVPQGVSHREAVAMVAPVTSAP	2628
Qy	1341	--CASDSPDRPNSTV-----ITTAN-----	1358
Db	2629	RPCMGPTPVQPSSELEPPEDPVOTQHLRILNPNVEYQODIPWRILFRKEVEFYPRKONSHP	2688
Qy	1359	-----RVLN-----CNADPREEHNNHITTLQSKSGDTRVBSGFIVRGWLHKRYKN	1404
Db	2689	VQDLDFRQILHDTFSEACRISDERLQMKALFAPONQDLTO-----RPLTESVKR	2740
Qy	1405	S-----PKMSL-----KLKRWFLV-----	1420
Db	2741	AALSMADSDMEIYFSRLFPAMGSVGTCVOGLAVSHSIGIKLQWVKGSKESRRRLVLCAY	2800
Qy	1421	-----THNSLDYKSSK-----NALKGLTV-----LNSLCSV	1450
Db	2801	SFADILEVTWPSQMLTEFINSNEKTLIFSARAQOVKTLVDTPIELKKSDDYVAVAENFL	2860
Qy	1451	PPDEKIRK-----ETGYWNTVYGRKICY-----	1474
Db	2861	SEDPBELLSEFKGDIILHQSLEPTRVGVSAGCVARKKLVYBELRRRGRDPFGMRGAVHGR	2920
Qy	1475	--RLYTKLNEAT-----RMSSVJONVDTKAPIDT-----PQOOLIO	1511
Db	2921	VGRPPESELVPRPAAPDFLOLPAERBGRRAAAVAAVASAAAAGVEGRRRBPARRASAD	2980
Qy	1512	IKENCTL--NSDVE--QIYKRPD-----ILRYTHNPLHSPL	1543
Db	2981	SGEDSIALPSTWMEFAQYKTFRDPRRRPRDGLKLKSKEDRESKTLBDVLCTFTKVPIOEST	3040
Qy	1544	LPLPYGDINANLKDQGYTTLQDEBAKIPNSLOO-----LEMSDPPIPIIOGLIOGCHD	1597

Db 3041 IEL--SDSLNKM-----AVDMFVAVRFNGDAPLKGQSE-LDVLCTLLKLCSD 3086
 Qy 1598 LRPLDELXCOLIKQTNKVPHPGSGVNLXSMOILCTLSCPLPSRGLKYLKRLKRIE 1657
 Db 3087 HEVWMEDECCQIVKQITDNSSPKQDSCQGMKLLIYMAVYGSEVFYLLRFLQHV-S 3145
 Qy 1658 QPPGTEMEKVALFTYESLKT--KREFVPSDEIEALI-HRQENTSTVYCHGGGCKI 1713
 Db 3146 WPPGJLPFGQIAKACEQNLQKTLRFGRLEFPSSNMEIRAMLAGRSSKROLSLPGILERL 3205
 Qy 1714 TINSHTAGENVKILRGILAME-----DSRNMALFENYCHDKALIESHTVAVDLAKE 1768
 Db 3206 KIKTCTVALDVIGELCTTEMALTRPEAFDEVIVFVNNRQCHV-CPUSCAVILDVASEME 3264
 Qy 1769 K-----LAATSEVCDLPKMKFYKLYCPLDTONVPDQSVFAFMFQAEAVHGHHPAP 1822
 Db 3265 QYDGGTLMFRVLMNQPLKFEHELIVTMHYNQVLPDYLGKF-----SSVPAQPF 3317
 Qy 1823 EENLOYLAALRYOYLOGDYTLAALPRLPEVYSLQRLKARISQSTYTFPCRLEKRTS 1882
 Db 3318 QOLOQVSKLASIQHRAKD--HFYLSVREVQ--EYIPAOVLYHTTAG----- 3359
 Qy 1883 FLEGTRRSFRGSSVVRKQVEEQMDMFIKEVSSARASIIDKMKFGGMNOEQAMAKY 1942
 Db 3360 -----DTWL-NLVSGHR-----OQTALSPHQRAQF 3385
 Qy 1943 MALIKEPGPGSTLPDVE-CKEGGPFQELMLGVSAVVS-----VYKR 1984
 Db 3386 LGLLSAFPFGSSFFIQCSCNVLPAPCLAVNHNGLMFLSKTHELVKIPLEKIQST 3445
 Qy 1985 GGGPRLPEVQYEH1-LSPGAPLANTYKIVDERELLFETSZVVDVAKMKAYISMVKKR 2043
 Db 3446 WTKQKTFANSYFVEISLD-----VAAQRTMQLQJELGRVAVVAHVESLMSAR 3497

RESULT 6
 ID MY15_HUMAN STANDARD; PRT; 3530 AA.
 AC 09UKN7;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Myosin XV (Unconventional myosin-15).
 GN MYO15A OR MYO15.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OC NCB1_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.
 RX MEDLINE=20021762; PubMed=10552926;
 RA Liang Y., Wang A., Belyantseva I.A., Anderson D.W., Probst F.J.,
 Barber T.D., Miller W., Touchman U.W., Jin L., Sullivan S.L.,
 Sellers J.R., Camper S.A., Lloyd R.V., Kachar B., Friedman T.B.,
 Friedell R.A.;
 RA "Characterization of the human and mouse unconventional myosin XV
 RT gene responsible for hereditary deafness DFNB3 and shaker 2.";
 RL Genomics 61:243-258(1999).
 RN 12
 RP PARTIAL SEQUENCE FROM N.A., AND VARIANTS DFNB3 TYR-2111 AND PHE-2113.
 RX MEDLINE=98267311; PubMed=9603736;
 RA Wang A., Liang Y., Friedell R.A., Probst F.J., Milcox E.R.,
 Touchman J.W., Morton C.C., Morell R.J., Noben-Trauth K., Camper S.A.,
 Friedman T.B.;
 RA "Association of unconventional myosin MYO15 mutations with human
 RT nonsyndromic deafness DFNB3.";
 RL Science 280:1447-1451(1998).
 CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
 CC ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
 CC COMPARTMENTS, WHICH WOULD BE MOVED RELATIVE TO ACTIN FILAMENTS (BY
 CC SIMILARITY). MAY PLAY A ROLE IN THE FORMATION OR MAINTENANCE OF
 CC THE ACTIN-RICH STRUCTURES OF THE INNER EAR SENSOR HAIR CELLS.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PITUITARY. ALSO EXPRESSED
 CC AT LOWER LEVELS IN ADULT BRAIN, KIDNEY, LIVER, LUNG, PANCREAS,
 CC PLACENTA AND SKELETAL MUSCLE. NOT EXPRESSED IN BRAIN. IN THE
 CC PITUITARY, HIGHLY EXPRESSED IN ANTERIOR GLAND CELLS.
 CC -1- DISEASE: Defects in MYO15A are the cause of an autosomal recessive
 CC form of nonsyndromic deafness (DFNB3).
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 IQ DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial/
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, AF144094; AAF05903.1; -.
 CC EMBL, AF051976; -; NOT_ANNOTATED_CDS.
 CC HSSP, P10587; 1BR2.
 CC Genew, HGNC:7594; MYO15A.
 CC MIM, 600316; -.
 CC InterPro: IPR000299; Band 4.1.
 CC InterPro: IPR000048; IQ region.
 CC InterPro: IPR000857; MYTH4.
 CC InterPro: IPR001452; SH3.
 CC InterPro: IPR001609; myosin head.
 CC Pfam: PF00063; myosin_head; 1.
 CC Pfam: PF00612; IQ; 3.
 CC Pfam: PF00784; MYTH4; 2.
 CC PRINTS: PR00193; MYOSINHEAVY.
 CC ProDom, PD000355; myosin_head; 1.
 CC SMART, SM00015; IQ; 2.
 CC SMART, SM00242; MYSC; 1.
 CC SMART, SM00139; MYTH4; 2.
 CC SMART, SM00326; SH3; 1.
 CC PROSITE, PS00660; BAND_41_1; FALSE_NEG.
 CC PROSITE, PS00661; BAND_41_2; FALSE_NEG.
 CC PROSITE, PS50057; BAND_41_3; 1.
 CC PROSITE, PS50096; IQ; 3.
 CC DR PROSITE, PS50002; SH3; FALSE_NEG.
 CC Myosin; ATP-binding; Actin-binding; Coiled coil; Repeat; SH3 domain;
 CC Calmodulin-binding; Disease mutation; Deafness.
 CC KW DOMAIN 1 1887
 CC FT DOMAIN 1888 2029
 CC FT DOMAIN 2030 3530
 CC FT DOMAIN 1323 1350
 CC FT DOMAIN 1792 1799
 CC FT DOMAIN 1902 1924
 CC FT DOMAIN 1925 1954
 CC FT DOMAIN 1955 1976
 CC FT DOMAIN 2867 2953
 CC FT DOMAIN 3206 3443
 CC FT NP_BIND 1315 1322
 CC FT VARIANT 2111 2111
 CC FT VARIANT 2113 2113
 CC FT VARIANT 2113 2113
 CC FT SEQUENCE 3530 AA; 395171 MW; 3D103523D48CBE4A CRC64;
 CC SQ
 CC Query Match 14.8%; Score 1589.5; DB 1; Length 3530;
 CC Best Local Similarity 25.1%; Pred. No. 4.6e-65;
 CC Matches 623; Conservative 348; Mismatches 818; Indels 697; Gaps 88;
 Qy 59 EGVGDMASTYELHGGSTIMVNLFGYKRNQIMWTYIGSTLASVNPQPIAGLYEPATMGY 118
 Db 1221 EDGVEDMQLQLEQLQTTVALSKIRFERNLLTYTGSLVSNPYQ-MGTYGPEQVOY 1279
 Qy 119 SRRHIGELPPIHFAIANBCYCLWRHNDQCILKIGESGAGKTBSTKLILKFLSVISDOOS 178

Db 1280 NGRALGPNPFLAVANLAFAKMLDAKONCIIISGSSGKTEATLILRLAAMOK- 1338
 Qy 179 LEISLEKTSQVBRALLESSPIMEAFGNATVYNNNSRFGKVOANTICQKNIQGRVY 238
 Db 1339 ----REVMOQIR--ILEATPLLESFGNATVRNDSRFGKVEFL--EGVIGALIT 1389
 Qy 239 DCILSSONRVARONPGERNTHIFVALLAGLEHEEREFEVYSTPENYHYNQSGVEDKTI 298
 Db 1390 SOYLEKSRIVFOAKNRNTHIFELLAGLPOLRQFSLQEAETYYLNOGNCELACK 1449
 Qy 299 SDOESFEVITAMDNQFSKEEVEVSRLLAGLILHGNIF-----ITAGAOVSFKTAL 353
 Db 1450 SDADDFRLLAAMEVLCFSSSEDODSIFRILASILHGNVFEKETEADQEAASVASAREI 1509
 Qy 354 GRASAILGLDPTQDALTQSRMFLNGEELITPLNQAADSDSLAMALYACCFFEMVIX 413
 Db 1510 QAVAEILQISPEGOKAITPKVTETMEKLTFTPLVESAVDARDALAKVYALLFSWLT 1569
 Qy 414 KINSRIKGNEDFKSIGLIDIFGFENFEVNHFEQFINIYANEKLOEYFNKHFSLQOLEYS 473
 Db 1570 RVAVASPRDITSLALDLYGFEDLSFNSFEOLCINANENIQLYFNKLVFOEQDEYI 1629
 Qy 474 REGIWEEDIMDNGECLDIE--KUGLALINEESHFOATDSTLLEKLHSHQANHFY 532
 Db 1630 REGIDMOEITFADNQPRIINLSKPYGILRLIDQCFCFQATDHTFLOKCHYHGANPLY 1689
 Qy 533 VKRNVANNGVHYAGEVOYDVRGILEKRDTFRODILNLRESRDFIYDF--EYVS 591
 Db 1690 SKRPMPEFTTIKHYAGKTYOVHKEFLDNDOVROVLDL FVRSRRVVAHFFSSIAPO 1749
 Qy 592 RNNQDTLKCS---KHRRPTVSSQFKVDSLHSLMATLSSNPFVRCIKENMOKMPDOF 648
 Db 1750 AAOQRCKSSSVRLYGAHTVAKFO--QSLDLVEKERCNPLFMRLKXNHKKEGJLE 1808
 Qy 649 QAVVLANOLRYSGMLETVIRIKAGVAVRRPODYKRYKVMRLALBEDVRGK---CTSL 705
 Db 1809 PDVVMQOLRYSGVLETVIRIKEGFVRLPFGCFIDRYCCL---VALGHDIPLANGDMCVSV 1865
 Qy 706 L--OLYDASNSMOGKTQVFLRESLBOKLEKREBEVSHAAVIRAHVGLFLARKOYRY 764
 Db 1866 LSLCLKVMYMYRGVSKLFLKEHLTYOLRESREHVNLAALTYOLRCLRFKIKRFRSL 1925
 Qy 765 LYCVIIIOKNYRAFLRLRRFLHLKKAIVFOKOLRGQIARVYROLIAEKREO--EEKKO 823
 Db 1926 RHHIILLQSRARGLARQYQOMRSLVKFRSLVHAVSRRLYKLABEWRQCVEGALLM 1985
 Qy 824 EEEBKKRR-----EEREREREREREAELRAQOEET 855
 Db 1986 EOEELSKREVVAAGHLEVPALAGLLOAVAGLQAVPQVAPVTRTPLQAEPRVTLPLDI 2045
 Qy 856 RKQOELEALQSKOE---AELTRELE-----KQEKQOYEBILREKELEIDQORKEQ 906
 Db 2046 NNYPMAPFVOCHFEPARFGMLTVPLTPTOLPAEHAAEAVSIFKILRLFMGDPHLHGAR 2105
 Qy 907 ELSLTEASLOK---LOERRDOELRLREBEACRAOEELSLNFDDEIDECVRIERS----- 959
 Db 2106 ENIFGNVTVQKGLAVPELRDEILQALANVH-----NHNANHERGMILL 2151
 Qy 960 ---LSGGSFSSSLASACEEKENFNFSQYPEEBEVDGFEADDDAFKOSPSEH--- 1012
 Db 2152 AACLISG-----FAPSPCNKYLKLFVSDYGR---NGFOA-----VCOHRLMG 2190
 Qy 1013 --GHSODRTSGITSDSSBEDPYMNDTVPTSPSADSTVLLAPSVODSG-----S 1061
 Db 2191 AMGRAPQOQSGGAAR-----TLPTQLEWTATYATEKASMALDVGCENGDSFSCP 2237
 Qy 1062 LHNSSSGESTYCMRONAGDLPSPDG-----DYDODDDED-- 1097
 Db 2238 VHSWSTGEV-----AGDILHRGLADGRGWTVAMKNGOVMAELAGHDYVLDVSDLE 2291
 Qy 1098 -----GATISGSSVTFSSNSYGSQWSPDYRCV----- 1124

Db 2292 LRLDEPRQKSYETVGTGEPAPASGRPKVFNGS-----WDSDEDMSTRPOEHEMKVLD 2347
 Qy 1125 ---GTVNSGAVRFSSEGAOSSFEDESED-----FDSRDTDELSTYRD 1166
 Db 2348 DGYSSHNOGCT--NGETEAORGTAHOESDLSCEPAVPHKGLDICYDLSLF--DPVLSYGA 2404
 Qy 1167 SVYSCVTLPEFHSFLYMKGLM-----NSWKRMCVLK-----DETFLMFRSKOBA 1212
 Db 2405 DEKPTALIV-----RMKGGQPGGSSSGCTEDTRRRPREKPIFGOLDASTL-----A 2452
 Qy 1213 LKQGLHKKGGSSSTLSRRNMKKRMFVLRQSL-----MYENDSEELKQTVERT-- 1264
 Db 2453 LOOAFIHXO---AVLAKGMTLOATLOOQPLSALSLSLPAEKPPAPAPAOPT--SVGTGP 2507
 Qy 1265 -AKEIIDNTT-----KENGIDIMADRTPHILASPEDASQ-----WF 1301
 Db 2508 PAKPVILRATPRPLAPAPAKAPRLPIKVAAPVLAQDA-----SPETTSPESELVRY 2561
 Qy 1302 SVLSQVH--ASTDOETQEM-----HDE----- 1321
 Db 2562 STLNSHFPQPTQIKNIYROYQOPFRGRPEALKKDGGKVFMKRPPDHEALMLIKGOM 2621
 Qy 1322 ---QANPONAAGTLDVGLDSCASDSP-----RPS 1351
 Db 2622 THLAAPGTOVAREVALVKPVTSAPREMAFTSALPSKSLBPBELTQTRLHLINPF 2681
 Qy 1352 F-----VIIVANVLH-----C----- 1363
 Db 2682 YGYQAPMKIFLRKEVFYKDSYSHPVQDLDLFRQILHDTLSBALCRISDEBLMKALF 2741
 Qy 1364 ---NADTPEEM-----HHWITLLO 1379
 Db 2742 AQONQDQPLVLTESVKRAVSTARDTWEVFSRIFPATGVSQVQLAVSHVGIKILR 2801
 Qy 1380 RSKGDTREVGQEFYIRGMLHKEVKNSPKMSLKLKRMVVL--TINSLDYYSSSEK---- 1433
 Db 2802 MYKGGQGAQGLRVLRVAFADI-----LEVTPSQMMEFLNLSKVLIF 2847
 Qy 1434 --NALKLGTV-----LNSLCSVVPDEKIF-----KETGY- 1462
 Db 2848 SARAHQVTLVDVDFILELKSDDYVAAYANFLPEBPALLAFHKDOIHLQPLEPRVGS 2907
 Qy 1463 -----MNV--TVYGRKHCYRLTYKLINAT----- 1485
 Db 2908 AGCVARAKVYVLELRRRGDPGMRFGTIGR--VGRFSEELVQPAAPDFQLPTEPGR 2965
 Qy 1486 -RMSYIOWNTYTKAPIDT-----PTQOLIQDIKENCL--NSDVVE--QIYKRP-- 1530
 Db 2966 GRAAAVAAAVASAAAQOEVGRRRBPVRARASADHGBALALPYTMEFAQKYFRDPGR 3025
 Qy 1531 -----ILRYTHNPLHPLLPYGDINMLKDKGYTLQDEAI 1569
 Db 3026 RPQDGLRLKSKEPRESRTLEDMLCTFTPLQSLLEL--SDSLSKMATDML-----AV 3078
 Qy 1570 KIFNSLOQLESNDPIPIIOGILQTHDLRLDELQCLIKQTKNVPHPSGVNLYSWO 1629
 Db 3079 MRFMGDAPLKGQSD--LDVLCNLKLKCGHEVWRDCYQVWQOQITDNTSSKQDSQORGR 3137
 Qy 1630 IITCL-----SCFTLSRGILKYLKHLKRIRE---QFPTBEMKALFYBELSKT--K 1679
 Db 3138 LLYIYVYAHXCS-----EVLHPLHTRFLQDVSRTPPLPQGIJAKACEOULQTLRFG 3189
 Qy 1680 CREFPSPRDEIALI--HROEMSTYVCHGGGCKTINSHTAGAVEKILGILAM--ED 1736
 Db 3190 GRLEIPSSIELPAMLAGSSKQQLRLRGLERELHKITKITCYALDVEIEICEMALTRE 3249
 Qy 1737 SHNMALPEYNG---HVDKALISRTVAVDLAKFER-----LAATSEVDLPMKFFYKL 1787
 Db 3250 AFNEVYIVFVTRRGHV--CPLSRAYIILDVASEMQVQGVLMFRVRVLMQDPLKEENEL 3308
 Qy 1788 YCFELTDNVFKQSVFAFMFEQAHAEVJHGHAPBEENI--QVLAALRQIYQSGYTLHA 1845
 Db 3309 YVTMYNQVLPDYLKLGFSSVPSR-----PSBOLLQOVSKLASLOHRAKD---HF 3356


```

Db      896 FRRUKAQAALKALKEIARSAEHLKRLVGMENKVVQLOQRKIDQNKEFKTLSEQLSAVTS 955
Qy      867 --SOKEMALRELEKOKENQVEEILLEKEIEDLOKMKQOQELSLTEALQLOQERDQ 924
Db      956 THANEVEKLKKEELARYOQNOBADPSLOOEVSOLR--TELQKHSRRVLEAHNNENG 1013
Qy      925 ELRR-----LEBACRAAQEFLESINPDEICVNIERSLSGSEFSESLAESACEKPN 980
Db      1014 ELKRVAIDLHEHNAIILKDE-KEHNLN-----HQILRQSAEBSGSSVBEHNLKKELEBRS 1068
Qy      981 --FNFQOPYPB-BEVDGEFADDDAFKOSPNSEHG--HSDQRTSGIRISDDSEDDPYM 1035
Db      1069 RYONLVEYEQLEBORYENLRDEOQTPGRKRNKPSNOSSLESDSVYPSISTEIDTEALQ 1128
Qy      1036 NDRTVPRPSGA-DSTVL-----APSVODSGSLH-----NSSGSEITCMQPNAGDLPS 1083
Db      1129 QVEEIGIEKAMDTVPLKLOKRVRELEQERKLOVLEKEQDDSKVQVEQONGEDVD 1188
Qy      1084 PDGDYDQDDYEDGATTSGSSVTFNSYSQOWSPDYRCVGTYNSSGAYRFSEGAQSS 1143
Db      1189 QADADIANSIKRQE--LESENKKLKNIDLERNK-----AVADQAMQDN 1229
Qy      1144 FEDSEEDFDSRF-----DTDELSYRDSVYSCVTLPEFHSFLYMKGLNWSMKRWCVL 1198
Db      1230 STHSPDSVSLNLNQLKLANEBLEVRKEEV-----LILRTQIMNADQRLSGK 1277
Qy      1199 KDETFL-----WFRSKEALKQGMWLNKKGGSSSTLSR-RWKKRWPLRQSKLMYENDS 1252
Db      1278 NMEPNARTSPWSEKHVDQEDALEYHGVCCQNSQTEBWG-----YLNEDG 1325
Qy      1253 E-----EKLKGVETRTAKEIINDTTKENGIDIMADRTPHLIAESPEDASQWF----- 1301
Db      1326 ELGLAAGLQVARNLEBAQOAGNLKHEEVENHKAQ--VEANKEMDKQOQFCQTILL 1383
Qy      1302 SVLSQVHASTDQEIQENHDEQANPQNAVGLD-----VGLIDSVCAQSDS 1345
Db      1384 SPAQVEFGVQOEISRLTNEIDLPEKLEKKEKKLQKQKIYMKVODLAAQALQ 1443
Qy      1346 PDRPNSVITITARVLHCNADTPEEMHMT-----LLORSQDPVREGEQETVAGWL--HK 1400
Db      1444 SDR-----RHHLTRQVTVORKSKD-----FGQMBLYHK 1472
Qy      1401 E-----VKN-----SPKSS-----LKLKRWPLTNSIDYKSSSEKNAIKGTLVANS 1445
Db      1473 EDEALLIRNLVTLOKPMISGTPCLPAYILLYMCIRHAD---YTNDLKHSHLSSTING 1529
Qy      1446 LGSVPPEDEKIFKETGYWNVTVYGRKCYRLYTKLNEATRWSSVIONVDTYAPIDTPT 1505
Db      1530 IKRYVLKKNHEDFEFTSFMLSNTRLLHLCLKQYSGDEGFMQ--WTAKQN----- 1576
Qy      1506 QQLIQIDIKENCL-NSDVEQIYKRNPIRLRTHHPLHSPILPLRYGVDINLNLMDKGYTLL 1564
Db      1577 -----BHCLKNFDLTE--YRQ-----VL 1592
Qy      1565 QDEAIKIFNSLOOL-ESMSDPIPI-----IOGILQTHDLRP----- 1600
Db      1593 SDLSIQVQOQLIKIABGLQPMIVSMLNESIQLSGVAPRTGYRRKSSSMVDGENSYCL 1652
Qy      1601 -----LRDELYCOLIKQTKNVPPHGVGNLYSMQILTCLSTPLPSR 1642
Db      1653 EAIIRQWNPFTVLCDQGLDPEIILQVFKOLFYMIVNAVTLNLN-----LRLKDCASW--ST 1706
Qy      1643 GILKYLKPHLKRIREQPGTEMKRY-ALFTYESL-----KTKCKEFPSPSDEITE 1691
Db      1707 GM--QLKYNISQLEEWLRGNLQOOSGAVQMEPLQAOLLOUKKKT-----QEDAE 1756
Qy      1692 AL-----IHRQMTSTVYCH--GGGSKITIN-SHTTAGEVEKILRGLAMEDSRNP 1741
Db      1757 AIGSLGTSLSSTQOIKLINLYTLPLNGFEERVTYSFRTIOAQOESNDPQOLLDSKHP 1816
Qy      1742 ALPEYN 1747

```

```

Db      1817 PVLFPFN 1823
RESULT 8
MS2_DICDI
ID MS2_DICDI STANDARD; PRT; 2116 AA.
AC P08799;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin II heavy chain, non muscle.
GN MYCA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxId=44689;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87092286; PubMed=3540939;
RA Warrick H.M., de Lozanne A., Leinwand L.A., Spudich J.A.;
RT "Conserved protein domains in a myosin heavy chain gene from
RL Dictyostelium discoideum.";
RN [2]
RP PHOSPHORYLATION SITES, AND MUTAGENESIS.
RX STRAIN=AX2;
RX MEDLINE=90353583; PubMed=2387408;
RA Lueck-Vielmeier D., Schleicher M., Grabatin B., Wippler J.,
RA Gerisch G.;
RT "Replacement of threonine residues by serine and alanine in a
RL phosphorylatable heavy chain fragment of Dictyostelium myosin II.";
RN [3]
RP PHOSPHORYLATION SITES.
RX MEDLINE=88112226; PubMed=2828113;
RA Wagler G., Noegel A., Scheel J., Gerisch G.;
RT "Phosphorylation of threonine residues on cloned fragments of the
RL Dictyostelium myosin heavy chain.";
RN [4]
RX X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 1-762.
RX MEDLINE=95345066; PubMed=7619795;
RA Fisher A.J., Smith C.A., Thoden J.B., Smith R., Sutok K., Holden H.M.,
RA Rayment I.;
RT "X-ray structures of the myosin motor domain of Dictyostelium
RL Biochemistry 34:8960-8972 (1995).
RN [5]
RX X-RAY CRYSTALLOGRAPHY (2.7 ANGSTROMS) OF 1-762.
RX MEDLINE=95345067; PubMed=7619796;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-pyrophosphate complex of the
RL truncated head of Dictyostelium discoideum myosin to 2.7-A
RN resolution.";
RN [6]
RX Biochemistry 34:8973-8981 (1995).
RN [7]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=96206189; PubMed=8611530;
RA Smith C.A., Rayment I.;
RT "X-ray structure of the magnesium(II)-ADP.vanadate complex of the
RL Dictyostelium discoideum myosin motor domain to 1.9-A resolution.";
RN [8]
RX Biochemistry 35:5404-5417 (1996).
RN [9]
RX X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-762.
RX MEDLINE=97452580; PubMed=9305951;
RA Gullick A.M., Bauer C.B., Thoden J.B., Rayment I.;
RT "X-ray structures of the MgADP, MgATPgammA, and MgAMPNP complexes
RL of the Dictyostelium discoideum myosin motor domain.";
RN [10]
RX Biochemistry 36:11619-11626 (1997).
RN [11]
RX X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 1-762.
RX MEDLINE=98070605; PubMed=9405148;
RA Bauer C.B., Kuhlman P.A., Bagehew C.R., Rayment I.;
RT "X-ray crystal structure and solution fluorescence characterization

```

RT of Mg.2'(3')-O-(N-methylanthraniloyle) nucleotides bound to the
 RT Dicyostellium discoideum myosin motor domain."
 RL J. Mol. Biol. 274:394-407(1997).
 CC - FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN & HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC - SUBUNIT: MYOSIN II HEAVY CHAIN IS TWO-HEADED. IT SELF-ASSEMBLES
 CC INTO FILAMENTS. HEXAMER OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI
 CC LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATORY LIGHT CHAIN SUBUNITS
 CC (MLC-2).
 CC - SUBCELLULAR LOCATION: HIGHEST CONCENTRATION IN THE POSTERIOR CELL
 CC CORTEX.
 CC - DOMAIN: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
 CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN BE FURTHER
 CC SPLIT INTO 2 GLOBAL SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
 CC SUBFRAGMENT (S2).
 CC - DOMAIN: THE ROD-LIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
 CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
 CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
 CC - PTM: PHOSPHORYLATION INHIBITS THICK FILAMENT FORMATION AND REDUCES
 CC THE ACTIN-ACTIVATED ATPASE ACTIVITY.
 CC - MISCELLANEOUS: DICTYOSTELIUM MYOSIN II HAS NO K(2) EDTA ATPASE
 CC ACTIVITY, PERHAPS CORRELATED WITH THE ABSENCE OF A CYS AT THE SH-1
 CC POSITION (688).
 CC - SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC - SIMILARITY: CONTAINS 1 IQ DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M14628; AAA33227.1; --
 DR PIR: A26555; A26555.
 DR PIR: S00250; S00250.
 DR PDB: 1MMA; 03-DEC-97.
 DR PDB: 1MMD; 17-AUG-96.
 DR PDB: 1MNG; 03-DEC-97.
 DR PDB: 1MNN; 03-DEC-97.
 DR PDB: 1MND; 17-AUG-96.
 DR PDB: 1MNE; 17-AUG-96.
 DR PDB: 1VOM; 23-DEC-96.
 DR PDB: 1LYK; 28-JAN-98.
 DR Dictydb: D001008; mhca.
 DR InterPro: IPR000048; IQ_region.
 DR InterPro: IPR004009; Myosin_N.
 DR InterPro: IPR001609; myosin_head.
 DR Pfam: PF00063; myosin_head. 1.
 DR Pfam: PF00612; IQ_2.
 DR Pfam: PF02736; Myosin_N_1.
 DR PRINTS: PR00193; MYOSINHEAVY.
 DR ProDom: PD000355; myosin_head. 1.
 DR SMART: SM00242; MYSC; 1.
 DR SMART: SM00242; MYSC; 1.
 DR PROSITE: PS50096; IQ_1.
 DR Myosin; Coiled coil; Actin-binding; ATP-binding; 3d-structure;
 DR Calmodulin-binding; Methylation; Alkylation; Phosphorylation;
 KW DOMAIN 1
 FT 762 791
 FT 817 2116
 FT NP BIND 179 186
 FT DOMAIN 638 660
 FT 738 752
 FT 130 130
 FT 678 678
 FT MOD RES 1823 1823
 FT MOD RES 1833 1833
 FT MOD RES 2029 2029
 FT SEQUENCE 2116 AA; 243871 MW; 2FC3770B1E556A1 CRC64;
 Query Match 14.0%; Score 1500; DB 1; Length 2116;

Best Local Similarity 35.3%; Pred. No. 3e-61;
 Matches 381; Conservative 220; Mismatches 355; Indels 124; Gaps 31;
 QY 6 VLRLRNGHPSTVNSCAGIVPRTDYGQVPTK-----QSTTHQKVTAMHPNTEGVS 62
 DB 35 IWTNPD---PERSYECGSISET--SDSFPTVDGQDQROVKDQDANRNPITKFGV 88
 QY 63 DDMASLTLEAGSGIMVNLFORYKRNQIWTYIGSILASVNPQIAGLYEPATMEQYSRRH 122
 DB 89 EDMSLSTLNEBAVPHNLRVAYNNDLIYTGSLPLVANNPKRIP-IYTOENVDFKRR 147
 QY 123 LGELPPIPAIANECYRCLMKRHDNOCILIGESGAGTSTKILKFLSVISQSLLS 182
 DB 148 RNEVAPHFAISDVAYRSMLDROQOSILLIGESGAGTENTKVIYVLASVGR----N 203
 QY 183 LKEKTSYCBRLABSSPIMEAFGNAKTYNNNSSRFGQVQNTIQKNNIGGRIVDCIL 242
 DB 204 QANGSGVLEQQLQNNPLLEAFGNKTRNNNSSRFGKFIQFNNAQPIGASIQSYLL 263
 QY 243 SSQNRVVRQNPGRNYHIFYALLAGLEHEEREPYSTPENHYLNGSGVEDKTSIDQ 302
 DB 264 -EKSNVYQSETERNYHIFYQLLAGATAEKKALHLAQPSPNYLNGGYVDIKVDSSE 322
 QY 303 SFREVITAMDVQPSKEEVRVSRLLAGILHGNIEPT-AG-GAQSFKTALGSAELL 360
 DB 323 EFKTRQAMDIVGFSQEQMSIFKIIAGILHGNIKFKGAGEGAVLKDXTALNAASTVF 382
 QY 361 GLDPQLDALTQRMFLRGEIILPLVNOQAVDSRDLAMALVACCEPWYIKTKNSIK 420
 DB 383 GNPSTLLEKALMEPILLAGRDVLAQHNVSSSRDLVALVAGRLPLMLVKTKNNVLC 442
 QY 421 GNEDKSGILDIFGFENFENVHFEQFNINYANEKLQGYFNKHFISLEQLSEYREGVME 480
 DB 443 QERKAYFGVLDISFELFKNSFQCLINTNEKQGFNNHMKLQDEVLKIKIWT 502
 QY 481 DIDW-IDNGECLDLEKKL--GLALLNEESHFPQATDSTLEKLSQHA--NNHFYVXR 536
 DB 503 FIDFELDSQATIDLDGRQPGIALLDEQSVFPAATNTLITKLHSFKSKNAKYBEP 562
 QY 537 VAVNFGVRYHAGVQYVVRGILEKRNPTFPDOLLNLRSEKRPDIYDLF--HYSSNN 594
 DB 563 FSKTEFGVTHAGQVWYIEIQMLEKNDKPLQDDLELCKDSSDNVYTKLPDNPNTASR-- 620
 QY 595 QDTLCKSGKRRPPTYSQFKYDSLHSLMATLSSSPFVRCIKPMKQKPPDQAVVLN 654
 DB 621 ---AKGKGNF--ITYAAQYK--EQLASLMAITLTPHVRCTIPNNKQDPAKLEKRVLD 674
 QY 655 QLRYSGLMELVIRKAGYAVRRPFQDFYKRYKVLNRNLAL--PEVYRGKCTSLQLYDASN 713
 DB 675 QLRGCVLEGRITRKFPNRIYADFVKRYVLLAPVPRDAEDSQKATDAVLKHLNDP 734
 QY 714 SEMQGLTKVPLRSLDEKLEKREBEVSHAAVYRAVHGLFARKYRKVLVCV---I 770
 DB 735 EQYRFGITKIFPRAQLRIEBAERQISEIKAIQAATRMIAKVKYQKAREHTVAARI 794
 QY 771 IQKNYRAF-----LRRR-----PLHLK----- 788
 DB 795 IQQNLRAIYDFKSNMWMKLFKAPRLKRPPEKIKEREILEKLSLNDSTTQDKL 854
 QY 789 -----KAALVPOKQDQG--QIARVYRQ--LAEKRQOE---EKKQEEBEKKRE 832
 DB 855 EKLSDTSNVLDLQRLKAKETLKAANYDSKDLBAQKRELEIVDEMESELDKLLAL 914
 QY 833 E--EREERERERAEMLAQQEERKQOELEALOKSOKAEALRELKQK--ENKQVEE 888
 DB 915 ETLQKQKSVKQVNDLEELQEQKLNTEKLKKYEE---ELBEMRVNDGQSDT 969
 QY 889 ILRLKEKIEDLQMKQOEQLSLTEASLQK-----LQRRDQELRLLEACRAQOE 939
 DB 970 ISRLKIKIDELQKVEELTESFSBSKQKGVLEKTRVLAQSELDDLTYRLDSET--KQKSE 1028
 QY 940 FLESINPEIDECVANIERSLSGSGSEFSSLAESAEEKPNFNSQPYPEBEVDEGFAD 999

Db 1029 LRFQKK--KLEBEIKQVQALAA--AETAAKLQAEANKKLQGEYV-----ELNKEFNSE 1078

RESULT 9

MY9B_HUMAN STANDARD; PRT; 2158 AA.

AC Q13459; Q9UHN0; O75314; Q9NUJ2;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DE MYO9B Ikb (Unconventional myosin-9b).

GN MYO9B OR MY9S.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.

OX NCBI_TaxID=9606;

RA [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver, and Small Intestine;

RX MEDLINE=97063843; PubMed=8907710;

RT Wirth J.A., Jensen K.A., Post P.L., Bement W.M., Mooseker M.S.;

RT "Human myosin-1xb, an unconventional myosin with a chimerin-like rho/rac GTPase-activating protein domain in its tail.";

RL J. Cell Sci. 109:653-661 (1996).

RN PARTIAL SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.

RP MEDLINE=20047919; PubMed=10580159;

RA Grewal P.K., Jones A.-M., Macconochie M., Lemmers R.J.F., Frants R.R., Hewitt J.E.;

RT "Cloning of the murine unconventional myosin gene Myo9b and identification of an alternative splicing.";

RL Gene 240:389-398 (1999).

RN [3]

RP SEQUENCE OF 1940-2158 FROM N.A. (LONG ISOFORM).

RX MEDLINE=98158729; PubMed=9490638;

RT Post P.L., Bokoch G.M., Mooseker M.S.;

RT "Human myosin-1xb is a mechanochemically active motor and a GAP for rho.";

RL J. Cell Sci. 111:941-950 (1998).

RN [4]

RP SEQUENCE OF 1828-2023 FROM N.A. (SHORT ISOFORM).

RC TISSUE=Placenta;

RA Itoigai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y., Nishikawa T., Nagai K., Sugano S., Ishibashi T., Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoaka S., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuko Y., Kaneshiro K.;

RA "NEDO human cDNA sequencing project.";

RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS. MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.

CC -1- SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE PERINUCLEAR REGION.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS, A LONG FORM (SHOWN HERE) AND A SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN PERIPHERAL BLOOD LEUCOCYTES AND AT LOWER LEVELS, IN THYMUS, SPLEEN, TESTIS, PROSTATE, OVARY, BRAIN, SMALL INTESTINE, AND LUNG.

CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG BINDING DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.

CC -1- CAUTION: THE C-TERMINAL SEQUENCE FROM AA 1917 ONWARDS FROM REF.1 WAS PROBABLY A CHIMERA.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation - CC the European Bioinformatics Institute. There are no restrictions on its CC use by non-profit institutions as long as its content is in no way CC modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (See <http://www.itsb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

CC EMBL; U42391; AAC50402.1; ALT SEQ.

CC EMBL; AF143684; AAF00119.1; -

CC EMBL; AF020267; AAC26597.1; -

CC EMBL; AK002201; BAA92132.1; ALT_INIT.

CC HSSP; P08799; 1MND.

CC Genew; HGNC:7609; MYO9B.

CC DR MIM; 602129; -

CC DR InterPro; IPR002219; DAG_PE-bind.

CC DR InterPro; IPR000048; IQ_region.

CC DR InterPro; IPR000159; RA_domain.

CC DR InterPro; IPR000198; RhogAP.

CC DR InterPro; IPR001609; myosin_head.

CC DR Pfam; PF00063; myosin_head; 2.

CC DR Pfam; PF00130; DAG_PE-bind; 1.

CC DR Pfam; PF00612; IQ_4.

CC DR Pfam; PF00620; RhogAP; 1.

CC DR Pfam; PF00788; RA; 1.

CC DR PRINTS; PR00193; MYOSINHEAVY.

CC DR ProDom; PD000355; myosin_head; 2.

CC DR SMART; SM00109; CI; 1.

CC DR SMART; SM00045; IQ; 3.

CC DR SMART; SM00242; MYSC; 1.

CC DR SMART; SM00314; RA; 1.

CC DR SMART; SM00324; RhogAP; 1.

CC DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.

CC DR PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.

CC DR PROSITE; PS50096; IQ_3.

CC KM Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;

CC KW Cytoskeleton; Coiled coil; GTPase activation; Phorbol-ester binding;

CC KX Zinc; Alternative splicing.

CC FT DOMAIN 1 939

CC FT DOMAIN 940 1044

CC FT DOMAIN 1045 2158

CC FT DOMAIN 1046 1071

CC FT DOMAIN 1880 1901

CC FT DOMAIN 1959 1989

CC FT DOMAIN 844 855

CC FT DOMAIN 957 977

CC FT DOMAIN 979 1000

CC FT DOMAIN 1001 1023

CC FT DOMAIN 1024 1053

CC FT DOMAIN 1633 1681

CC FT DOMAIN 1713 1862

CC FT NP_BIND 239 246

CC FT VARSPPLIC 2022 2023

CC FT VARSPPLIC 2024 2158

CC FT CONFLICT 1937 1939

CC FT CONFLICT 1947 1947

CC FT CONFLICT 2040 2045

CC FT CONFLICT 2049 2049

CC FT CONFLICT 2067 2067

CC FT CONFLICT 2157 2158

CC SQ SEQUENCE 2158 AA; 243556 MW; 4978F1D70F56D28 CRC64;

Query Match 13.9%; Score 1492; DB 1; Length 2158;

Best Local Similarity 25.5%; Pred. No. 7,1e-61;

Matches 498; Conservative 329; Mismatches 701; Indels 428; Gaps 60;

QY 11 RENGHPSTVNSCAEGIVFRIDYGVFTYKOSTITTHQK-TAMHPNBSGVDDMASLT 69

DB 101 QEDGYFFLQERN-ADGTI---KYVHQLVAQATATRLVERGLPQQADFDLCNLP 155

QY 70 ELHGSIWVNFQYKRNQIWTYGISLASVNPQPIAGIVPATMEQYSRHGELPPH 129

DB 156 ELTGNLKLNLKRLRLOOKITVYAGSILVAINPK-FLPTNPKYVKNYENQGLKLEPH 214

QY 130 IFAIANECYCLMKRDNOCILIKGSSGAKTESTLILKFLSVISQOLESLEKRETC 189
 DB 215 VALALNAVAYTMLRKVNOCIVYPGSSGGKTOSTFLIE:CLTALSO-----KCYASG 267
 QY 190 VERRAILESSPIWEAFNAKTVVNNSSRGKPFQULNICOKGNIOGGRIVDCLTSSONRV 249
 DB 268 VERTIIIGACPVELEAFNAKTAHANNNSSRFKPIQVSYLESGLVRGA--VEKILLESKRLV 326
 QY 250 RONPGHRYHIFAYLAGLEHEHEEEFYLTSTPENHYALNCGCEDKTSIDOE---SPR 305
 DB 327 SOEKORHRYHVFYLLGLVSEBERGFOLKOPEDYTYLNC---HNLKIEDGDLKHDE 382
 QY 306 EYITANDVNOFSKEEYREVSRLLAGILHAGNEFTAG---GAOVSKTALGRSAELL 360
 DB 383 RIKQANEMVGFIPATKQIFAVYLSAILYGNVYKRRATGRBEGLEVPPEVLDLSQL 442
 QY 361 GLDPTQDALTQSRMFLRGEELTPLNQAANDSDSLMALYACCFPMVYKTIKSRJK 420
 DB 443 KVRREILVELTGRKTVTVNDKILIPYSLEAITADSMASKSYSLFPMIVLRINHALL 502
 QY 421 GNEDFK-----SIGLIDIFGFENFEVNHPEOFNINANEKLOEYFNKHFSLQLEGR 474
 DB 503 NKKDVEANVSCISIGVDFGPEDFERNSEFOFCIYANEOLOYVYNOHIFKLEGEYOG 562
 QY 475 EGLWEDIDWIDNGECLDLIEKK-LGLLALINEESHFPOATDSTILLEKLSQHANNHYV 533
 DB 563 EGIYTHNIGYTDVNGCIIHLSKPKTGLFYLLDEESNFPHTSQTLLAKFQOJEDKRYFL 622
 QY 534 KRRVANNGVGVYAGEVOYDVRGILEKRDTRDILLNRSRDFIYDLF----- 586
 DB 623 GTPVNEPAPFIQHPACKVKYQIKDFREKMDYMRPDIALLRGSSDYVRELIGMDPVAV 682
 QY 587 -----BH----- 588
 DB 683 FRMAVILRAIRAMAVLREAGRLAERAEKAGSSGSAOSHPELPRGASTSEKLYRDL 742
 QY 589 ----- 588
 DB 743 HNQMIKSIGLPMWGEDPRSLQSLRLOKPRAFILSKGIKQOII PKNLLDSKSLKLI 802
 QY 589 -----VSSRNNOPTLKCGSKHRRPTVSSOPKVDLSHLMATLSSNFFPRCTIKPNQKRP 644
 DB 803 ISMTLDRITKSLHLHKKKPPSISAQFOT-SLNTLLLEALGAEPPFRICRSNAEKE 861
 QY 645 DQPDQAVVNLQSLYSGLMETVIRIRKAGAVRARPQDFYKRYKVMNLALPEDV---RQK 701
 DB 862 LCFDEDLVIOQLRYTGMLETVARIRSGYSAKTFQDFEQFVL-----LPKDAQCREV 916
 QY 702 CTSLLQLYDASNSEWQGTQVFLRESLQKLEKREEEVSHAAVIRAHVGLFLARKQY 761
 DB 917 ISTLEKMKIDIKRNVIGIKTKVFLKETERQALQETLHREVVRKILLQSMFVRLERHF 976
 QY 762 RYLYVCVYIIQKNYAFPLRRRFLHLKKAIVFOKOLGOIARVYR-----QLIA 812
 DB 977 LQMKRAAVTIOACWRSYVRRA-LERTQAAVYLOAAMRGVORKLRYHOKOSIIRLOSJC 1035
 QY 813 EKRBOEKKKOESEKKRESEERER-ERERREAEILRAOEBETTRQOELALQSKXA 871
 DB 1036 RGHLOKFSQMSISEKQAKBEKRELELARAAGAECSGCGDAAAGQOAAQGEPEBDG 1095
 QY 872 BLTRELEKQENKQVEILRLKEIEDLQRMKEQOELSTLEASLOQERRODELARLEE 931
 DB 1096 HLASBEVQPSDRSPLHSSPEKEAVSPEKTLRPOQTVAASHHEKYPSSREKRESNR--- 1152
 QY 932 EACRAAOEFLSINFP--EIDECVNRISLSGGS--EFSSELAESECEKXPNFNSQYR 988
 DB 1153 -----QRLHVKFKQNKHIOQCKE--ESALRPSRRVVOBOGVSLLEK-----K 1195
 QY 989 BEEVEDECF-----EADDDAFK--DSPNPEHSHSDORTS3ITSDSDSEBDPMYMDTV 1040
 DB 1196 ESEMEDTLVLEAENTSOKEPTQEPQMAVAKVSEETB--KTLPSGS---PRQOLER 1250
 QY 1041 PTPSPADSTVLLAPSVQDSGLHNSSGSESTYCMPONAGDLP--SPGDYDYDDQDYEDG- 1098

DB 1251 PTLSLDRLV-----SPAPGASPEPPEDEKSKCGSPRVOEKDSDSGSTQIOIR--YLDAB 1304
 QY 1099 -----ATSGSSYFSSYSGOWSPDYRCISVGTNSGAVRFSRGAOSSPEDE--- 1148
 DB 1305 RLASAVELMRGKVLAAASPSAMLSQSLDLS--DRIRATGALLTPBERITSSTDVSKL 1363
 QY 1149 -----EDPDSREPTDELSTYRDSYSCVTLPEFHSFLYMKGLMNSMKRWCVLKDET 1203
 DB 1364 LPSLAKQAPAAETTDGESSAKKPAVOK----- 1390
 QY 1204 LMFRSKQALKOGMLHKKGGGSSITL-----SRNWKRMFVLROSKLMTFEND 1251
 DB 1391 -----KKRGDASSILPDAGLSPGSOVDSKSTPFKR--LFLHKTK-----D 1426
 QY 1252 SEEXLKGTVRYRTAKE---IIDNTKENGIDIMADRPHILAE---SPEDASQWPSVL 1304
 DB 1427 KKYSLGABELENAVSGHVVLEAITMKGLEPSCQQRHAAAGEKRTKEPGGKKKRRNV 1486
 QY 1305 SOVHASTDOELQEMHDEQANPONAVGTLDVGLIDVSCASDSDPRPNSFVITIANRVLHCN 1364
 DB 1487 KIGKITVEKMRRESVFRQITANLKLIDEFLNKNINDLRSGKPTIESLFIATSKPRSN 1546
 QY 1365 ADT-----PEBMHHITLL-----QSKGDTREVGQEF-----YRGMLHK 1400
 DB 1547 IKTMYSVNGKIHGVYKDLMEYQIVVSNLATERGQKDTNVLNLFQSLDDEFTRGYTKN 1606
 QY 1401 E---VNSPKNSSLSLKKRWFLTNSS---LDYSSSKNMLK-----LGTVLNLSL 1447
 DB 1607 DEEPVKQS-KQKQKRRKORAAVEHNGHVPASVOYSIOSECQCLSYIMLMDKALLCSVC 1665
 QY 1448 SVVPDEKIPKETYGMANTVYGRKHCYRLYTGLLEATRMSVIONVTDKRAPIDTPQO 1507
 DB 1666 KTMCHKCVHKIQSHCSYTL-YGRK-----GEPGAEHGFVCVDSLSDKASVPIVLKX 1718
 QY 1508 LIQDIKENCLSNVVEQIYKEN-----PIIRYTHPLHS----- 1541
 DB 1719 LLEHVEHMGH---YTEGLYRKSGANRTRELRQALQTDPAVKLENPFHAIITGVIAKWL 1775
 QY 1542 -----PLPL-PLYGINNLKDKGTYTLODEAKIFINSLOQESMSDPIPIQOIILOTG 1595
 DB 1776 RELPEPLTFAYQGFPL-----RAVLEPKQEQALAI---YAALEHLEPKN 1818
 QY 1596 HDL--RPLRDELYCQLIKQTNKVPHPGSGVNLYSQIOLTC--LSCTPLPSRGLLYLKFH 1651
 DB 1819 HNSLERLLFHLVKVALLDVAIRM--SPGALAIIFAPCLRLCPNDSPLSMKDVAKITTCV 1877
 QY 1652 LKRIREQPPTGEMEKYALFTYESLKTTCREFPVSR 1687
 DB 1878 EWLKEQ-----MRKYKV-KMBEISQLEMAESIARF 1907
 RESULT 10
 MYSJ DICDI
 AC P54697;
 ID MYSJ DICDI STANDARD; PRT; 2245 AA.
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin II heavy chain.
 GN MYOJ.
 OS Dictyostelium discoideum (slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelidae; Dictyostelium.
 OX NCBI_Taxid:44689;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX3;
 RX MEDLINE=96215148; PubMed=8636147;
 RA Hammer J.A. III, Jung G.;
 RT "The sequence of the dictyostelium myo J heavy chain gene predicts a
 RT novel, dimeric, unconventional myosin with a heavy chain molecular
 RT mass of 258 kDa."
 RL J. Biol. Chem. 271:7120-7127 (1996).

RN [2]
 RP SEQUENCE OF 1-1021 FROM N.A.
 RX MEDLINE=97039016; PubMed=884597;
 RA Peterson M.D., Uricole A.S., Titus M.A.;
 RT "Dictyostelium discoideum myoI: a member of a broadly defined myosin
 v class or a class XI unconventional myosin?";
 RL J. Molec. Res. Cell Motil. 17:411-424(1996).
 RN [3]
 RP SEQUENCE OF 182-298 FROM N.A.
 RX MEDLINE=95023928; PubMed=7937787;
 RA Titus M.A., Kuropa A., Loomis W.F.;
 RT "Discovery of myosin genes by physical mapping in Dictyostelium";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).
 CC -1- FUNCTION: MYOSIN IS A PROTEIN THAT BINDS TO ACTIN AND HAS ATPASE
 CC ACTIVITY THAT IS ACTIVATED BY ACTIN.
 CC -1- SUBUNIT: HOMODIMER.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 3 IO DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: UA2409; AAA5186.1; -;
 CC EMBL: L35322; AAA79858.1; -;
 CC HSP: P08799; 1MND.
 CC DictyDb: DD01095; myoJ.
 CC InterPro: IPR002710; DIL.
 CC InterPro: IPR000048; IO region.
 CC InterPro: IPR004009; Myosin_N.
 CC InterPro: IPR001609; myosin_head.
 CC Pfam: PF00063; myosin_head; 2.
 CC Pfam: PF00612; IO; 6.
 CC Pfam: PF01843; DIL; 1.
 CC Pfam: PF02736; Myosin_N; 1.
 CC PRINTS: PRO0193; MYOSINHEAVY.
 CC ProDom: PD000355; myosin_head; 2.
 CC ProDom: PD003376; DIL; 1.
 CC SMART: SM00015; IO; 3.
 CC SMART: SM00242; MYSC; 1.
 CC PROSITE: PSS0096; IO; 3.
 CC Myosin, ATP-binding; Actin-binding; Calmodulin-binding; Repeat;
 KM Coiled coil.
 FT DOMAIN 1 809 MYOSIN HEAD-LIKE.
 FT 824 851 IO 1.
 FT DOMAIN 872 901 IO 2.
 FT DOMAIN 943 972 IO 3.
 FT DOMAIN 973 1812 COILED COIL (POTENTIAL).
 FT DOMAIN 1813 2245 TAIL.
 FT NP_BIND 174 181 ATP (POTENTIAL).
 FT DOMAIN 669 749 ACTIN-BINDING.
 FT CONFICT 191 191 L -> F (IN REF. 2).
 FT CONFICT 284 284 A -> T (IN REF. 2).
 FT CONFICT 291 291 G -> R (IN REF. 2).
 FT CONFICT 332 347 NKGCEIEGVSDDEH -> IEMPELKYVRMS (IN
 FT REF. 2).
 FT CONFICT 550 550 N -> K (IN REF. 2).
 FT CONFICT 865 866 HH -> QQ (IN REF. 2).
 SO SEQUENCE 2245 AA; 258478 MW; 615E5EF1D1AB55B CRC64;
 Query Match 13.3%; Score 1490.5; DB 1; Length 2245;
 Best Local Similarity 24.5%; Pred. No. 8.8e-61;
 Matches 529; Conservative 377; Mismatches 665; Indels 589; Gaps 71;
 QY 4 OGTRWLRN-----GHPSTVNSCAEGIVFRTYGVFTKOSTITHQXTANHPN 58
 DB 26 EGAGVWIPDELMTGIGADVIEHSETSDOVLVRTEDREV-----KIPLSKVFOKPNPI 79

QY 59 EGVDDMASLTELHGSGIMVNLFORYKKNQIWTYIGSILASVNPQIAGLYEPATMEQY 118
 DB 80 LGVDDLSFLSHLHPALAHNLHRYNLNQIYTIKILIAINPTSLP-LYKEMISAY 138
 QY 119 SRRLGELRPPIFALANECYRCMLMKRD--NCCILKESGNGKTESTLLIKFSLVISO 176
 DB 139 YKQIGTALPHYAVVAEDAFKDM--RYDTSQSLVSGSGAGKETTYTFLQYLAAMGN 196
 QY 177 QSLLELSEKETS-----CVERALLESSPIMEAFGN 206
 DB 197 M-----IKSTSSSSINGINTSSDGIPTPPSPMKKSRYDVSVERVLESTPLEAFGN 251
 QY 207 AKTVNNNSNRFGKVLQNI CQKNIQGRIYDCLSSONRVNRQNGERNHYIFALIA 266
 DB 252 AKTLRNDNSRFGKFLIEHFNMGSILGAKILAYLL-EGSIVRQVYERNHYIFQYLLS 310
 QY 267 GLEHERREPFYLTENHYILNOSGVEKETSIDQSPREVTANDVQFSSEYREVER 326
 DB 311 GASELKEKELNKTIEEYSLNKSQCFIEGVSDDEHFNKTCMAQVAGITLVEQENVR 370
 QY 327 LLAGLILHGNIEFTAGG-----QVSPKALGRSAELLG-LDPTQLTALQRSWFLRG 380
 DB 371 ILSAILINFEFENIAGSNDSCOLIDRPLEKYSVLGCAQDPBELNSMLTRKYVTK 430
 QY 381 EEILPPLVQAQVDSRDSIAMLVYACCEEWIKIKINSRIKGNEDFS--IGILDFGE 437
 DB 431 ESYISHTKERAENNRDLSMFLYGMFDMVLVKINSMSISTQKSKSFGIVLDIYGE 490
 QY 438 NFEVNHFEOPNINYNANEKLOEYFNKHIPSLBELYRSBELVVEDIDMIDNGCLDLEK 497
 DB 491 SEFVNGFEQFCINYNNEKLOLFNQHVFEEQOEYIKEDIMSYIDFNQDTLLEIKN 550
 QY 498 -LGLTALINBSHPQATDSTLLEKLSQHANHHYVVRVANFVGHYAGEVOYDVR 556
 DB 551 PICILTLDEITMFKAIPQTLATLYSKMTHSKFEKRPSTAFTHNHVAGKTYEYD 610
 QY 557 GILEKXNDTFRDLLNLRESRFPIYDLFEH-----VSSRNNQDTLK 599
 DB 611 QFLDKNKDPIIEQISLIORSNFSFIKVLMSHSDKFTQSPGHPQNGGPTSSNTKGT-S 669
 QY 600 CGSKRRPPIVSSQFVDSILHSLMATLSSNPFVNCICPMQKMDQFQDAVVLQKRS 659
 DB 670 GSSSMKFLSVSGQFST-SLATLMKTIITTPHYVVCIKENPKLPQTEFKQOVHQLRG 728
 QY 660 GMLETVRIKAGVAVRRPDPFYKRYKVLN-----RNLLPEYVRGKSTL 705
 DB 729 GWESVRIICAFPRRRLISEFYQRYKILYVDINTGSGGKSGSNKKIKPKLVLQNL 788
 QY 706 LQLYDANSSEWQGLTKVFLRESLEQLEKREBEVSHAAYIRAHVIGFLARKQYRY- 764
 DB 789 LIGIELSDDKYKIGLTKVFLRAGQASLEDMLEQLDRSATVIOKRMKGYLYRKYKQLR 848
 QY 765 -----LYCVIITQKYNRAF-----LLRR 782
 DB 849 DASLIITQKRSVNAKHLISALQRTHTSAILIQKVRFAHRDRVOYOKIRDASIQLOTWNR 908
 QY 783 RFL-----HKKAIVFO-----KOLRG-----QIARVYROLIA 812
 DB 909 HLFSEQVHREKENAIIITQIKRIQLSKREVDKLRGILLIQAMRMKLARGVITQLRA 968
 QY 813 EKREGEKKKQEEBEKKREBEERERERERERERERERERERERERERERERERER 871
 DB 969 EAR-----SLRTVQEQNKQLEKLELQWLTSEA 998
 QY 872 ELTRELEKX-----ENKQVE-EILRLKEKLELDLQRMEOQLSTESLQVLT 918
 DB 999 KKKQQLQEDQKVSDDTTIELSSNNDBLLEQLSEIQLOEIDNSQSSQLQSEC-LSLTL 1057
 QY 919 QERDQ-----ELRLSEPAQRAQEFLESLNFDIECVRIIEISLGSSEFSESLAES 973
 DB 1058 EEQTQQLDHSSTLANKLEKLDSDQHSIEKLO-SQPNTEIQQLQFQKQSELSKTLST 1116
 QY 974 ACEERPNFNSQPYDEEVEDEGFADDDAFKDSPNPSEHSGSDORTSGIRTSDDSEBDP 1033


```

Db      1117 T--QQLDFP-----KQEPD-----RISQERDNDNTN----- 1140
Qy      1034 YWMDTVVPTSPSADSTVLLAPSVQDSGLHNSSGSESTCMQPNAGDLLSPGDYDYPD 1093
Db      1141 --NQLFIQQLKKAQNSTL----- 1157
Qy      1094 DYEDGATITSGSVTFSNSYGSQWSPDRCSVGTYNSSGAFRPSSEGAQSPEDSEDPDS 1153
Db      1158 DY-----PQLSGIRNMLE--RQVLEL 1176
Qy      1154 RPDTDDELSTYRDSVYSCVTLTYFHSFLYMKGGLNMSWFRNCVLDKDETFMFRSQEAL 1213
Db      1177 R-DENQLIKERLDSLQO-----OSSQFQSGAALKEQOLQIV-QEJSEQLIKSSEKL 1227
Qy      1214 KQGMHLKKGSGSSTLSRRWKKRWFPLRQSKLMPENDSEKLLKGVN----- 1261
Db      1228 GSEERAKKQINOLELELTDHKSQI--OLQLTEGSENEKIKLKGLESTQDEKQOLOE 1285
Qy      1262 --VTAKEIIDNTTKENGIDIIIMDRTH--LIASEPCLASQWPSVLQVHASTDOEIQ 1316
Db      1266 LERIKSQSQSVED--EKNSLITQLTVKPESTQVSTNVSHQEKITLTLSTBELNKSIG 1343
Qy      1317 EKHDEQANFQNAVGLDGLID-----SVCSADSEDPNPGSVITANRVLHC 1363
Db      1344 KLOAQCKKNDDEIRIKIOPELNDQKQOFTROTKEPSPDLOSCOSIDRKESEITI-----HS 1397
Qy      1364 NADTEEMHMTTLLOBSKGDTRVEGOEP--IVRGMLHKEYVNSPKHSLKCKRWFL 1420
Db      1398 LERTMETLKSDFERVOQSLKQOERDQOQKDTINR--LENEVA-----OLQOKEF--- 1447
Qy      1421 THNSLDYKXSEKNAKLGTVLNLSLCSVVPPEKIFKETGYWNVTVGRKHCYRLYKL 1480
Db      1448 ---ENEFFVAKQNSNQ-----TQESVYLK-----EYLTQWQNSRIERL 1486
Qy      1461 LNEATRWSSVIONVDTKAPITPTQQLIQDIKENCLNSDVVEQIYKRNPIIRYTHHPLH 1540
Db      1487 BEKKOHITRIDDERBELKKQLTQLOQHQSSTQLLALQNEERLRKKE--LKY----- 1538
Qy      1541 SPLPLPYGDIMLNLKDKGYTT-----QODEAKIFNSLQOLESSSDPIPIIGIL 1592
Db      1539 -----KENGHERSKQOQDPNMEIQSLRTTN--DQLKSLQD----- 1572
Qy      1593 QTHGDLRLPLRDELYC--QLIKOTNKVPHSGVNLVSMOILTCLSTFLPSRGILKYLKF 1650
Db      1573 -YEQEKKKLKDLSKSKQEAQQRE-----SIHKMAELSAIQ 1610
Qy      1651 HLKRIREFPGTEM-----EKYALFTYESIKTKCREPVPSPDEIS-ALIHQEMTS 1701
Db      1611 HQQWENSEF--TDMKQPNQELIESSALYKQQLQOSTSTIDSTIKKEKENEISKLOQOLETS 1668
Qy      1702 TYVCGGSCKITINSHTTAGEVVELINGLAMEDSRMNFALFEVNGHDKALIESTVVA 1761
Db      1669 NOQLH---OLKEELNSMKOSNOLE-----STBOSKQNLQLOEQOLK----- 1708
Qy      1762 DVLAKFEKLAATSEVQDLPMKFEYFKLYCFLTDVNPVKDSVEFAFMFEQAHAVIHGHHPA 1821
Db      1709 -----SYTNEIS-----KQDDAVF----- 1723
Qy      1822 PSENLOVLAALQLYIQG-----DYTLHAIPLEEVYSLJRIKARISOSTTPTPCERL 1876
Db      1724 --ENOKINNTIKEOETKSRMSVELQOHIDEGKQOEIQOLQJSTIAQLKQOQOSET--DRL 1779
Qy      1877 EKRRTSPLEGTLRSPRTSGSVVRQKVEEOMLDMWKE--EYSSARASIIIO---KMK 1929
Db      1780 EKEIQO-----MKREKET-----OMKLVESTKLVNHLMEJRMELYRNVMELIDYKETEWEK 1830

```

```

RESULT 11
MY9B_RAT
ID MY9B_RAT STANDARD; PRT, 1980 AA.
AC 06358;
DT 16-OCT-2001 (rel. 40, Created)

```

```

DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Myosin IXb (Unconventional myosin-9b).
GN      MYO9B OR MYR5.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxId=10116;
RN      (1)
RP      SEQUENCE FROM N. A.
RC      STRAIN=Sprague-Dawley; TISSUE=Brain stem, and Spinal cord;
RX      MEDLINE=9518874; PubMed=7882973;
RA      Reihard J., Scheel A.A., Diekmann D., Hall A., Ruppert C.,
RA      Baehler M.;
RT      "A novel type of myosin implicated in signalling by rho family
RT      GTPases.";
RL      EMBO J. 14:697-704(1995).
CC      - FUNCTION: MYOSINS ARE ACTIN-BASED MOTOR MOLECULES WITH ATPASE
CC      ACTIVITY. UNCONVENTIONAL MYOSINS SERVE IN INTRACELLULAR MOVEMENTS.
CC      MAY BE INVOLVED IN THE REMODELING OF THE ACTIN CYTOSKELETON. BINDS
CC      ACTIN WITH HIGH AFFINITY BOTH IN THE ABSENCE AND PRESENCE OF ATP
CC      AND ITS MECHANOCHEMICAL ACTIVITY IS INHIBITED BY CALCIUM IONS.
CC      ALSO ACTS AS A GTPASE ACTIVATING PROTEIN ON RHO.
CC      - SUBCELLULAR LOCATION: IN UNDIFFERENTIATED CELLS COLOCALIZES WITH
CC      F-ACTIN IN THE CELL PERIPHERY WHILE IN DIFFERENTIATED CELLS ITS
CC      LOCALIZATION IS CYTOPLASMIC WITH THE HIGHEST LEVELS IN THE
CC      PERINUCLEAR REGION (BY SIMILARITY).
CC      - TISSUE SPECIFICITY: EXPRESSED IN TESTIS, LUNG, THYMUS, BRAIN,
CC      LIVER, AND SPLEEN.
CC      - SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHOSPHO-ESTER AND DAG
CC      BINDING DOMAIN.
CC      - SIMILARITY: CONTAINS 1 RHO-GAP DOMAIN.
CC      - SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC      - SIMILARITY: CONTAINS 4 IQ DOMAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL, X77609; CAA54700.1; -.
DR      HSSP, P08799; 1MND.
DR      InterPro, IPR002219; DAG_PE-bind.
DR      InterPro, IPR000048; IQ_region.
DR      InterPro, IPR000159; RA_domain.
DR      InterPro, IPR000198; RhoGAP.
DR      InterPro, IPR001609; myosin_head.
DR      Pfam, PF00063; myosin_head_2.
DR      Pfam, PF00130; DAG_PE-bind; 1.
DR      Pfam, PF00612; IQ_4.
DR      Pfam, PF00620; RhoGAP; 1.
DR      PRINTS, PR00193; MYOSINHEAVY.
DR      ProDom, PD000355; myosin_head; 2.
DR      SMART, SM00109; C1; 1.
DR      SMART, SM00015; IQ; 3.
DR      SMART, SM00242; MYSC; 1.
DR      SMART, SM00314; RA; 1.
DR      SMART, SM00324; RhoGAP; 1.
DR      PROSITE, PS00479; DAG_PE_BIND_DOM_1; 1.
DR      PROSITE, PS50081; DAG_PE_BIND_DOM_2; 1.
DR      PROSITE, PS50096; IQ_3.
KW      Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
KW      Cytoskeleton; Coiled coil; GTPase activation; Phospho-ester binding;
KW      Zinc.
FT      DOMAIN 1 940 HEAD OR MOTOR DOMAIN.
FT      DOMAIN 941 1045 NECK OR REGULATORY DOMAIN.
FT      DOMAIN 1046 1980 TAIL.
FT      DOMAIN 1841 1861 COILED COIL (POTENTIAL).
FT      DOMAIN 1918 1948 COILED COIL (POTENTIAL).

```

FT DOMAIN 845 856 ACTIN-BINDING.
 FT DOMAIN 978 978 IO 1.
 FT DOMAIN 981 1001 IO 2.
 FT DOMAIN 1002 1024 IO 3.
 FT DOMAIN 1025 1054 IO 4.
 FT DOMAIN 1593 1641 PHORBOL-ESTER AND DAG BINDING.
 FT DOMAIN 1673 1822 RHO-GAP.
 NP BIND 239 246 ATP (POTENTIAL).
 SQ SEQUENCE 1980 AA; 225035 MW; D79FECAD0FA8C05 CRC64;

Query Match 13.7%; Score 1471; DB 1; Length 1980;
 Best Local Similarity 25.3%; Pident. No. 5,86-60;
 Matches 497; Conservative 309; Mismatches 670; Indels 492; Gaps 60;

11 RENGHPSTVNSCAEIGVFRFDYGVFTYKOSTITTHOKV-TAMHPTEEGVDMASTL 69
 101 RENGYYFLDQERN-ADGSI-----QYLHVQLLAQPTACRLVEGGLPRPADDDLCNLP 155
 70 ELHGSGIMVNLFORYKRNQIMTYIGSILASVNPYQIAGLYEPATWEOYSRRHIGELPPH 129
 156 ELNEANMLQSLKLPVQOKIYTAGSILVAINEFK-FLPIYNPKYVMENQOLGKLEPH 214
 130 IPIANECRCILKMRDNOICILKSGSGAKTESTKILKFLSVISQOSELSEKERTSC 189
 215 VFPLADAVAYTMLRKHYNOICIVISGSGSGKTOSTNFLHCLTALSO-----KGYASG 267
 190 VERALIESPIMEAFGNAKTVYNNNSRFGKVFQNLICQKNTIGGRIVDCILSONRVY 249
 268 VERTIIGAGFVLEAFGNAKTAAHNNNSRFGKFIQVNTYLSGIVRGA-VVEKYLLEKSRIV 326
 250 RQNGERNYHIFYALLAGLEHEBEREPYLTSPENNYHLNQGCEVEDKITSDOE----SPR 305
 327 SQEKDERNYHVFYLLGVSEBERQFOLKQPODYFLNQ-----HNILNEDGEDLKHDEP 382
 306 EVTTANDWQFSKEEVRVSRLLAGLILHCLNIEFTAG-----GAQVSEPTALGRABELL 360
 383 RLQQAEMVGFPLATKQKQIFSVLSAILYGNVYTKRATGRDGLGVGPREVDLTLSQL 442
 361 GLDPTOLTALTORSMFLRGEILITPLNQAVDSRLMALYACCFEVIKINSRIK 420
 443 KYARELLEVLTRKITTIVNDKILIPYSLEAITAADSMAKSLYSALFDMIVIRINHAL 502
 421 GNEDFK-----SIGLIDIFGFENFEVNEQFNINANEKLOEYFNKHFISLEQLEYSR 474
 503 NKMDMEAVSCISIGVLDIFGFEDFERNSPEFCINAVNEQLOYFTQHFILKEQEBYGO 562
 475 EGIWMDIMDINGECLDLIEK-LGLLALINEESHFPQATDSTLEKLSQAHNHFY 533
 563 EGISWHNIDYTDVNGCIIHLISKRPGLFYLLDESNFPHATSHTLAKFQOHEDNKYFL 622
 534 KPIVAVNNEGVKHYAGEVOYDVAGILEKNDTFRDLLNLRESRPDIYDLF----- 586
 623 GPFLVEPAFLIQHFAKRVKQIKDPRKNDYMRPDIVALLRGSDSDSYARQLIGMPVAV 682
 587 -----EHVS----- 590
 683 FRWAVLRAAIRAMAVLREAGRLRAERAKEAGVSPVTRSHVEELPRGANTPSEKLYRD 742
 591 ----- 590
 743 LHNQIIKSLKGLWGEDPRLQLSLQKPRTPFLKSGIKQIKQOIIIPKNLDSKSLRL 802
 591 -----SRNNQDTLCKGSHRRPTVSSOPKVDLSLMAATLSSNPFVRCIKPMQOK 643
 803 IISMTLHDKTTKSLHLHKKKKPPSISAQPT-SLKKLLEALKAEPFIRCTIRSAEKK 861
 644 PQDFDAVVLNQLRYSQMLETVIRIRAGYAVRRPQDFYKRYKVLNRLNALPEDV--RG 700
 862 ELCFDEDELVLQGRYTGMLETVIRIRSGYSAKYTFQDFTEQFQVL-----LPDVOGRCE 916
 701 KCTSLIQLIDANSQWQGTQKTFVLESLEQLEKREBEV----- 741
 917 AIAALLEKLOVDNRQNYIGKTKVFLKETERQALQERLHGEVLARRILLQSWFRMVLRRH 976

742 ----SHAAMVIRAHVGLFARKQYRKVLVCVIIQKNYRAFLRRRFLHLKKAIVFOKO 797
 977 FVQMKHAALTIQACMSRYVRRTLERT-RAAVYLLQAAKRGVQLQRAHYHQRISIRLQSL 1035
 798 LRQIARRYROLAEKQOEKKQOEBEKKRE-----EERERERERER- 845
 1036 CRGHLQRSFSQMMLEKQAEQARETAGAEMSEGPSYVAAGEQSEHPVEDPESIGVET 1095
 846 -----ELBAQOEERTRKQOELEALQKSOKAEALRELEKOKENQOVBET 889
 1096 EYWMNSKSPNGISPKPEI-PSPEMETPAQKIVPA--ESHKVPSSREKRESRRQGLEIV 1152
 890 LRLEKEIEDLQMKQOEELSTLASLOKLOERRDDELRLREBACRAAQEFLSINPDEI 949
 1153 EKQNHQISCR-----BENSTLRPRSKASLETGES--PPE- 1186
 950 DECVNRNERSLSGSGSEFSELAEASACEKXPNFSCQPIPEEVDGCFEADDDAFDQSPNP 1009
 1187 -----DTKEPRREDGLETWETETAPSCPKQVPIVGDP-----RSBSP 1223
 1010 SEHGHS-----DORTSGIRTS-----DSSSEDPYNNQTVV--PTSPSADS-----TVL 1051
 1224 LQRPASLDLSKVSFVLPSSSLESPODEKGENSTKVODKPSFGSGTQIQRYQHPDTER 1283
 1052 LAPSYOD-----SGSLHNSSGSESTYCMFQNAAGDLPSPDGDYDQDDYEDGAI 1100
 1284 LATAVIEMRGKTLASAMLSQSLDSEK-----PRTAGAALT----- 1320
 1101 TSGSSVTFSNYSYGSMSPDYRCVSTYNSGAYRPSSEGAQSSFPDSEDPSPRDTDE 1160
 1321 TEERISFSTSDVSKLSP--VKISTEVDGDLSAKKRPGHKKKSDPSGAPAGLPTGSQ 1377
 1161 LSYRRDSVSCVTLTPYHPSFLMKGLMNSMKRRCVCLKDEFLPMRSGOZALKQGMHLK 1220
 1378 -----GDSKSAFKR-----LFLHAKDKKPSLGEVET 1405
 1221 KCGGSSTLSSRRMKRWFLRQSKLMEFNSEEEKLKG-----TYEVRTAKEII 1269
 1406 EBSGQQAQEAARATLDV--PSSQOHNHTTEKRLKQKNNRNVGQITVSEKKRESVF 1463
 1270 DNTYKENGIDITMADRTPHLIAESPEDASQMFVLSQVHASTDQEIQEMHDEQANPQNAV 1329
 1464 RKITYANAEKFL-----DEFILNKVNDLRSOKPIESLFEATERPRNSMTWYSVFN-- 1516
 1330 GTLDVGLDSVCASDPRPNSFVITANRVL-HCNAPTEEMHMITLLQ--RSKQDT 1385
 1517 GKIHVGKDLM-----ENYQIVSISLAERGEKOTNLVLANFQSLDEFTSYNKT 1567
 1386 RVEQGEFIVRGWLAKEVKNSPKMSLSLKRWFLVLTNLSLDYKSSERNAALK----- 1437
 1566 DEEP-----VKG-----KAQKKRKQBEAVALGHNH-NVFAASYOVNVPQSCQCLSY 1612
 1438 ---LGTVLNLSLCSVPPDEKTFKETGYMNTVYRKHCYRLYKTLNLEATMSSVIONV 1494
 1613 IWLMDKALLCSVCNKTCHKKCVHKIQSYCSYT--GRRK-----SELGAEPGFHGVCSVDL 1665
 1495 TDTKAPIDPTQQLQDIKENCLNSDVQIYKRPILRYTHRPHSPULPYPGDINLN 1554
 1666 TSDKASVPFLVEKLELHEVMBGL--YTBGLFRKSGAANRTR-----ELKOA 1709
 1555 LKDKGYTTLQDEALIKIFNSL--QQLSEMSDPIPIQOILQTHDLR-----PLRDE----- 1604
 1710 LQTDATAYKLEDFPIHATITGVKQMLRELREP--LMPFAQVGDPLRAVELBEKEQDLAA 1766
 1605 LYC-----OLIKQTNKVPHPQSGVNLYSWQIITC--LSCTFL 1639
 1767 IYAVLDHLPEANHTSLERLIFHLVYKALLLEDVNRH--SEGALAIIPAPCILRPRDSDPLT 1825
 1640 PSRGILKYLKFLKRIEQPFTGEMEKALFTYESLKKTKCEPFPVR 1687
 1826 SMKDLVKITTVCEMLIKQ-----MRKYV-KMEINHLAEASIAFR 1867

RESULT 12
 MYSA_HUMAN STANDARD; PRT: 1855 AA.
 AC 09YAT1: Q9UE30, Q9UE31, Q07902, Q16249, Q06653;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle)
 GN (Myosin heavy chain 12) (Myosin).
 OS MYOSA OR MYH12.
 OC Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
 RC TISSUE=skin;
 RA Meurers B.H., Zimmermann R., Vosberg H.P.;
 RT "The complete cDNA for human myosin heavy chain 12, a class V
 RT myosin.";
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT CYS-1246.
 RX MEDLINE=9751514; PubMed=9207796;
 RA Pascual E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
 RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
 RL Nat. Genet. 23:373-373 (1999).
 RN [4]
 RP SEQUENCE OF 638-1477 FROM N.A. (ISOFORM 2).
 RC TISSUE=fetal brain;
 RX MEDLINE=94245227; PubMed=8188282;
 RA Engle L.J., Kennett R.H.;
 RT "Cloning, analysis, and chromosomal localization of myosin (MYH12),
 RT the human homologue to the mouse dilute gene.";
 RL Genomics 19:407-416 (1994).
 RN [5]
 RP SEQUENCE OF 1061-1498 FROM N.A. (ISOFORM 2).
 RC TISSUE=brain;
 RX MEDLINE=95136715; PubMed=7835087;
 RA Moore K.J., Testa J.R., Francke U., Mlatovich A., Copeland N.G.,
 RA Jenkins N.A.;
 RT "Cloning and regional assignment of the human myosin heavy chain 12
 RT (MYH12) gene to chromosome band 15q21.";
 RL Cytogenet. Cell Genet. 69:53-58 (1995).
 RN [6]
 RP SEQUENCE OF 1358-1460 FROM N.A. (ISOFORM 3).
 RA Edgar A.J., Bennett J.P.;
 RT "Inhibition of dendrite formation in melanocytes transiently
 RT transfected with antisense DNA to myosin V.";
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP FUNCTION.
 RX MEDLINE=99376094; PubMed=10448864;
 RA Mehta A.D., Rock R.S., Rief M., Spudich J.A., Mooseker M.S.,
 RA Cheney R.E.;
 RT "Myosin-V is a processive actin-based motor.";
 RL Nature 400:590-593 (1999).
 CC -1- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
 CC STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
 CC MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
 CC BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
 CC FORMATION.
 CC -1- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
 CC CALMODULIN OR MYOSIN LIGHT CHAINS.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE

CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DISEASE: DEFECTS IN MYOSA ARE A CAUSE OF GRISCELLI SYNDROME (GS).
 CC GS IS A RARE AUTOSOMAL RECESSIVE DISORDER THAT RESULTS IN
 CC PIGMENTARY DILUTION OF THE SKIN AND HAIR, THE PRESENCE OF LARGE
 CC CLUMPS OF PIGMENT IN HAIR SHAFTS, AND AN ACCUMULATION OF
 CC MELANOSOMES IN MELANOCYTES. MOST PATIENTS ALSO DEVELOP AN
 CC UNCONTROLLED T LYMPHOCYTE AND MACROPHAGE ACTIVATION SYNDROME,
 CC KNOWN AS HEMOPHAGOCYTIC SYNDROME, LEADING TO DEATH IN THE ABSENCE
 CC OF BONE MARROW TRANSPLANTATION.
 CC -1- SIMILARITY: CONTRAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTRAINS 6 IQ DOMAINS.
 CC -1- SIMILARITY: CONTRAINS 1 DILUTE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL, U90942; AAD00702.1; -;
 CC DR EMBL, Y07759; CAA69035.1; -;
 CC DR EMBL, Y07759; CAA69036.1; -;
 CC DR EMBL, 222957; CAA80533.1; -;
 CC DR EMBL, S74799; AAB33211.1; -;
 CC DR EMBL, AF055459; AAC14188.1; -;
 CC DR HSSP, P10587; 1BR2.
 CC DR Genew; HGNC:7602; MYOSA.
 CC DR MIM, 160777; -;
 CC DR MIM, 214450; -;
 CC DR InterPro; IPR002710; DIL.
 CC DR InterPro; IPR000048; IQ_region.
 CC DR InterPro; IPR001609; myosin_head.
 CC DR Pfam; PF00063; myosin_head; 1.
 CC DR Pfam; PF00612; IQ; 6.
 CC DR Pfam; PF01843; DIL; 1.
 CC DR PRINTS; PR00193; MYOSINHEAVY.
 CC DR Prodom; PD000355; myosin_head; 1.
 CC DR Prodom; PD003376; DIL; 1.
 CC DR SMART; SM00015; IQ; 6.
 CC DR SMART; SM00242; MYSC; 1.
 CC DR PROSITE; PS50096; IQ; 6.
 CC DR Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
 CC Coiled coil; Phosphorylation; Alternative splicing; Polymorphism.
 CC KW Coiled coil; Phosphorylation; Alternative splicing; Polymorphism.
 CC FT DOMAIN 1 765
 CC FT DOMAIN 766 788
 CC FT DOMAIN 789 818
 CC FT DOMAIN 814 836
 CC FT DOMAIN 837 861
 CC FT DOMAIN 862 883
 CC FT DOMAIN 885 914
 CC FT DOMAIN 914 1237
 CC FT DOMAIN 1237 1445
 CC FT NP_BIND 1687 1792
 CC FT MOD_RES 163 170
 CC FT MOD_RES 643 665
 CC FT VARSPPLIC 1321 1347
 CC FT VARSPPLIC 1413 1413
 CC FT VARIANT 1246 1246
 CC FT CONFLICT 668 668
 CC FT CONFLICT 833 833
 CC FT CONFLICT 863 863
 CC FT CONFLICT 922 922
 CC FT CONFLICT 1061 1061
 CC FT CONFLICT 1089 1089
 CC FT CONFLICT 1177 1177
 CC FT CONFLICT 1465 1477
 CC FT CONFLICT 1471 1471
 CC /FTID=VAR_010645.
 CC F-> L (IN REF. 1).
 CC MISSING (IN REF. 4).
 CC E-> G (IN REF. 1).
 CC H-> R (IN REF. 1).
 CC V-> L (IN REF. 5).
 CC E-> Q (IN REF. 4).
 CC D-> E (IN REF. 5).
 CC NIPREKPOGML-> SVLCACCVSVTVR (IN REF.
 CC 4).
 CC K-> N (IN REF. 5).

FT CONFLICT 1484 1484 E -> D (IN REF. 5).
 SQ SEQUENCE 1855 AA, 215419 MM, 1C55AD57285FA9EC CRC64;
 Query Match 13.7%; Score 1466; DB 1; Length 1855;
 Best Local Similarity 25.7%; Pred. No. 9.1e-60;
 Matches 517; Conservative 349; Mismatches 681; Indels 468; Gaps 78;

QY 7 RWLRENGOHFSTVNSCAB-----GIWPRDYGVFTYKQSTTHQVTAMHPT 57
 DB 12 RWIIPD-----PEWVMSAELLDKDYKGDVLLHLHEEGKDLHLDPKTEK.PHLRNP 66
 QY 58 NEGVDMASTLHGSSIMYN.FQYKRNQ-ITWTIGSLAVNPQIPLAGIYEPATME 116
 DB 67 ILVGENDLTALSYLHBPVAVLHNRVRFIDSKLIYTCGIVLVAINPEQOLP-IYGEDJIN 125
 QY 117 QYSRHRLGELPHIFAIANECYRCMLWRHNOICLLKGSAGKSTSTKLLKFLVSIQ 176
 DB 126 AYSQNMGDNDPHIFAVAEAYKQMADEKNOIIVSGSGAGKTVSAKAMRYFATVSG 185
 QY 177 QSLLESLKEKTSCEVERAILESSPIMEAFNAKTVNNSSRFGKFVOLNICOXNIQGR 236
 DB 186 SASEANVEEK-----VLTSNPIMESIGNAKTTRBNDNSRFGKYIEIGFDKRYRIIGAN 238
 QY 237 IYUCIISQNRVVRQNGERNYHIFALLAGLEHBEREEFYLTSPENYHILNOSGCVEDK 296
 DB 239 MRYTL-EXRSVVFQAEERNYHIFQOLCASAKLPFKMLRLGNADNFNTYTKQSGSPVIE 297
 QY 297 TISDOSEFREVITAMDVQFSKEEVSRLLAGLILGNIEFTTAGQAVSF---KTL 353
 DB 298 GVDADKEMATRAQCTILIGISESHQMGIFRILNGILHLNGVGTSRDSDCTIPKHEPL 357
 QY 354 GRSAGELGDPOTLTALQSRNFEELRGEELTPLNVOAVDSRSLMALVYACCFEMVIK 413
 DB 358 CIFCDLNGVDYEBEHCMLCHRLATATETETIKPISKLOATNARDALAKHAYALKFWIYD 417
 QY 414 KINSRI-KGNBEDKSGIILDIPEFENFEVNHFEQFNINANEKLOEYFNHIFSLQLEY 472
 DB 418 NVAQALHASAVKQHSFGLVDIYGETFEFINSFOFINANEXKLOOFNNHNVKLEOBEY 477
 QY 473 SREGLVEMEDIDWIDNEECIDLIEKKLGLALINEESHFPOATDSTLLEKLSOHANN-HF 531
 DB 478 MKQOITWTLIDPFDNOPCINLIESKGIILDLDEECMFGKTDITMAQKYNHNLKCAL 537
 QY 532 YKPRVAVNFGVYHAGEVQYDVRGILEKRDTPFDLNLNRESRFPYIDLFEHVS 591
 DB 538 FEKPRLSNKAFFIQHFADKVEYQCEGFLKKNKDTVEBOIKVLSKSFKMLPELF----- 592
 QY 592 RNNODPLKGS-----KRRPTVSSQPKVDLSHSLMA 623
 DB 593 ---QDEKALISPTSATSSGRTPLTRTPAKETKGRPGQMAKHNKKTGVGHQPR-NSHLHME 648
 QY 624 TSSSNPFVRCIKPNQKMPDOAVLVNOLRYSGMLETIVIRAGYAVRPRFQDFYK 683
 DB 649 TLMATTPHYKCIKPNDFKFPFTFEKRAVOQLRACGVLETTIRISANGPFSRTTYOFS 708
 QY 684 RYKVLMENTALPEVDGKSTSLQLVDASNEWQKTKVFLRESLEOKLEKREBEVSH 743
 DB 709 RYKVLKQKQVLDSDRKQTCNVLEKILLDKQYQFGTKIFPRAGQVAVLEKLRADKLA 768
 QY 744 AANVTIAHVGLFARQYKRVLCV-----ITQKYRAP 778
 DB 769 ACIRIKTIRGWLIRKKYLMRAATMORVYRGYARCYAKFLRTKATIIIOKRWY 828
 QY 779 LARRRFLHKKAIIVOKOLQIARVRQOLAE-----KXQOE 819
 DB 829 VARRKIRRAATIVQSYLRGLARRRKIRKILREKHAIVIQKRVGMLARTHYKSMHA 888
 QY 820 -----KKQOEKKEKKEEERER-----ERREALRAQOE----- 853
 DB 889 IIVLOCFRMAKRELKUKIARSAVERYKHLIGEMKNIMQLOKRVDEQNDYKCI 948
 QY 854 -----ETRK-QOELEALQSKQEAELT--RELEKQENKQVEELIRLEKEIEDU 899

DB 949 KLTNLEGINSTETKLSOLERLQSEBEAKVATGRLVSLQ-----EBIANLRKDL- 1000
 QY 900 QRMKEQOELSTEASLOKQERDOELRLJEEACRAAOEFLESIN---FDEIDECVRNI 956
 DB 1001 ---QTRSEKCCIEBHADRKYQETEOVLSUKENTLLKQE-KEALNHRIVQAKETTEM 1056
 QY 957 ERSLSGSEFSELASACEKPNF---NFSQYPEEVEDEGFADDAFKDSPNPE 1011
 DB 1057 EKKL---VEETQLELDLNDERLRYQNLNFSR--LEERYD-LKEEWTLMVHPK-- 1108
 QY 1012 HGHSDORTSGITSDS-----SEEDPYNDTVPPSPASDSTVLAPV----- 1056
 DB 1109 -GH-KRTDSTHSSNESEYIFSSSEIAEMEDIPSRTEEPEKKVPIDMSLFLKQRYTEL 1165
 QY 1057 -ODSGSLHSSSGSESTYCPQNAAGLPSPD---GDYD---ODYEDGATITSGSYTF 1108
 DB 1166 BEQVQMDDELKREQVLRSAKKEERQIRGAELEYSLKQLEESNKKLKNELNL 1225
 QY 1109 SNSYGSQWSPDYRCVGYNNSGAYRFSSEGAQSFSEDEEDFSRFTDDELRYRDSV 1168
 DB 1226 RALSEKSAPEV---TAPGAPAYRVLMEQLTS-----VSEELDVRKEEV 1266
 QY 1169 YSCVTLPRHSLFYKKGGLMSW-----KRRVCYLKDETFM-----FRSQEALQOGL 1218
 DB 1267 -----LILRSQVSKQKALIQPDKNMTDSTILLEVDQMKDKGE-IAQAYI 1313
 QY 1219 H-KKGGSSSTLSRRMW---KRMFYL-----QSKLMYFENDSEKLGTYE 1261
 DB 1314 GUKETRSSALDYHELNEDGELMYEGKQANRLLESQLOQKXSHEDA-EALRG-E 1370
 QY 1262 VRTAKEIIDNTKENGIDITIMADRTPHLAE--PEDASQWFSVLSQVHASTDOETQEM 1318
 DB 1371 IOSLKE-----EN-----NROOQLAQMLOLPEARIBASIQHETRLTNELDLM 1416
 QY 1319 HDEQANPQ-----AVGTLVDLISVCS---DSPDRNSVITIANV 1360
 DB 1417 --EQLKQDQTVRKLLQKQVPAKKIGIELEVQOMENISPGQIIDEPTR----- 1463
 QY 1361 LHGNADTPREMHMTITLQSRKQDTRVEGEQFIVRGMLHKEVNSPKMSLKLKR----- 1416
 DB 1464 ---VNIRKEDPQGMLEYKEDQ-----KLVN---LILELKRGVAV 1502
 QY 1417 -----WFLTHNSLDYKSSSEKNAKLGTLVNLSCSVVPPDEKIFKETGYMV 1465
 DB 1503 NLPGLPAVILPWCVRH--ADVLNDQK-VRSLLTSTINSIKKVLKRGDDPEVSPWLS 1559
 QY 1466 TYVGRKHCYRLTK---LNEATRMSSVYQVNTOTKAPIDPFTQOIIDIKENCIN--- 1518
 DB 1560 NTCRFLHCLKQYSGEGBFKHNTSR---QN-----EHCULTND 1594
 QY 1519 -----SDVVEQIYKRNPIRTHHPHSLPLP--LPYGD1-NILNLKDKGY---TT 1563
 DB 1595 LAEYQVSLDLAQIYQO--LVRLVENILOPMTIVSGMLEHETIQGVSGVKPGLAKRIS 1652
 QY 1564 LODDAIKLFNS-LQOLESMSDPIPIIOGILQTHDLRLPRLDELYQLIKQTNKVPHPGSV 1622
 DB 1653 IADEGYTLDSILROLNFSH-----VMCOHQMDP--ELIKQVWQOMYIIGAITL 1701
 QY 1623 GNLVSMQILTCSCFPLBRSGLKYLKFLHKLIRQFPQTE-MEXKALFTYESL----- 1675
 DB 1702>NNL---LIRKDMCSW--SKGN--QIRYVNSOLEBMLDKNIMNSGAETLEPLLOAOL 1753
 QY 1676 ---KTKCRFEFVPARDEIALIHRQEMTSYVYCHGSGSCKI--TINSHTTAGEVEK- 1728
 DB 1754 LOVKKTD-----DDEALCS-----MCAALTAQIVKVLNLATTPNBEVERVS 1797
 QY 1729 ---IRGLAME-----DSRNPAL-FEYN 1747
 DB 1798 VSFIRTIQRLADBRDPSQLMDAKHIFVTEPPFN 1832

RESULT 13.
 MYSC_HUMAN

QY 1126 TYNSSGAYRPSSEGAOSSFEDSEBDFSDRPTDDE---LSYRDSVYSCTVLPYFHSFL 1181
 1130 HLNEDGELMFAVGLKATRVLSHPSQDCYEKEIEALNFK--VHLSQELNHQKFL 1187
 QY 1182 YMGGLMNSKKRKCVCULKDTFLMFRKQKALKQGLHKKGGSSSTLSRNNKKRMFVLR 1241
 DB 1188 REENDINESIHEVTRLTSENMMI PDKQO-----ISELEKPKODLEIR 1231
 QY 1242 QSKLAFENDSEELKCTVE-----VTAKEIIDNTKENGIDILMA 1283
 DB 1232 L-----NEQAEKMKKLEELSNQHLRSEBECTQKALEQNEIHTKEKELDKICE 1284
 QY 1284 DRTFHLIAESPEDASQMFVSLQVHASTDGEIQEMHDEQAN-----PQNAV 1329
 DB 1285 -----MQFASDHLKQKFETSESEKCFRQGEASRLTLENDLBEELDMKQRIKQLQDPV 1338
 QY 1330 GTLD--VGLIDSCASSPRPMSFVITANRVLHCNADLPBEMHMTITLQSKGPTRY 1387
 DB 1339 KTLSTKTIKANDVHSSSGP-----KEYLGMLOKREDEAK 1373
 QY 1388 EGOEFIVRGWLHKEVKNKSPMSLKLKRMFV-----LTNNSLDYKS-SEKNAL 1436
 DB 1374 LIONLI-----LDLKKRGVYVNMIPGLPAHLFMCVRYADSLNDANML 1416
 QY 1437 K-LGTLVNSLCSVPPDEKIFRETGYMNTVYGRKHCYRLYTKLNEATRMSSVIONVY 1495
 DB 1417 KSLMNSTINGIKQVVKHEHDFEMLSPWLSNTHPLNCLKQVSG-ESEFPMKNSPQON-- 1473
 QY 1496 DTAAPIDPTQOQLIDIKENCLN-----SDVEQIYKRNPLRTRH-HPLHSLP 1543
 DB 1474 -----KNCLNPNFLSEYRQILSDVARIYHOFIIMEKNLOPTIVPG 1515
 QY 1544 LPLPYGDI-NLNLTKDKGY---TTLQD-EATKIFNSLOOLEMSDPPIIQLITQTHD 1597
 DB 1516 M-LEYSLOISGLKPRGFRKSSIDDTGTYMTSVLOLST-----FTTMCQNGLD 1568
 QY 1598 LRLRLDELYCOLIKQ 1612
 DB 1569 -----PELVQRAVKQ 1578

RESULT 14
 MYSA_RAT
 ID MYSA_RAT STANDARD; PRT; 1828 AA.
 AC Q90YF3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle).
 GN MYO5A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathia; Muridae; Rattus.
 NC NCBI_Taxid:10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20380823; Pubmed=10920234;
 RA Futaki S., Takagiuchi Y., Hayashi Y., Ohmori S., Kanou Y., Inoue M.,
 RA Oda S., Seo H., Iwakawa Y., Murata Y.,
 RT "Identification of a novel myosin-Va mutation in an ataxic mutant
 rat, dilute-opisthotonus.";
 RL Mamm. Genome 11:649-655(2000).
 CC -1- FUNCTION: PROCESSION ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
 CC STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
 CC MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
 CC BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
 CC FORMATION (BY SIMILARITY).
 CC -1- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
 CC CALMODULIN OR MYOSIN LIGHT CHAINS.
 CC -1- DISEASE: Defects in MYO5A are a cause of Dilute-opisthotonus
 CC (dop). Dop rate have diluted coat color and are occasionally
 CC associated with severe neurological disorders.

CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 6 IQ DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB035736; BAA88350.1; -;
 CC HSSP: P10587; 1BR2.
 CC InterPro: IPR002710; DIL.
 CC InterPro: IPR000048; IQ_region.
 CC InterPro: IPR001609; myosin_head.
 CC Pfam: PF00063; myosin_head; 1.
 CC Pfam: PF00612; IQ; 6.
 CC Pfam: PF01843; DIL; 1.
 CC PRINTS: PR00193; MYOSINHEAVY.
 CC ProDom: PD000355; myosin_head; 1.
 CC ProDom: PD003376; DIL; 1.
 CC SMART: SM00015; IQ; 6.
 CC SMART: SM00242; MYSC; 1.
 CC PROSITE: PS50096; IQ; 6.
 CC Myosin; Repeat; ATP-binding; Calmodulin-binding; Actin-binding;
 CC Coiled coil; Phosphorylation.
 CC KW COILED COIL; MYOSIN HEAD-LIKE.
 CC FT DOMAIN 1 765
 CC FT DOMAIN 2 766 788
 CC FT DOMAIN 3 789 813
 CC FT DOMAIN 4 814 836
 CC FT DOMAIN 5 837 861
 CC FT DOMAIN 6 862 884
 CC FT DOMAIN 7 885 914
 CC FT DOMAIN 8 914 1239
 CC FT DOMAIN 9 1314 1418
 CC FT DOMAIN 10 1660 1765
 CC FT NP BIND 163 170
 CC FT MOD RES 643 665
 CC FT MOD RES 1733 1733
 CC SQ SEQUENCE 1828 AA; 211762 MW; 5B3DE1C89AB36123 CRC64;
 CC -----
 CC Query Match 13.5%; Score 1444.5; DB 1; Length 1828;
 CC Best Local Similarity 25.6%; Pred. No. 8.6e-59;
 CC Matches 511; Conservative 345; Mismatches 683; Indels 457; Gaps 74;
 CC -----
 CC 7 RWLLENGQHFPSTYNSCAE-----GIYFRIDYGVQVFTYKOSTITTHQKRAMPT 57
 CC 12 RWIIPD-----PEEWKSAELLDKYPGDKVLLHLEBKDLERYLDPRTSELPHLRND 66
 CC 58 NEEGVDMASLTELGGSIMYNLFORYKNO-IMTYIGSILASVNPYOPRAGLYEPATME 116
 CC 67 ILVENDLTAALSYLHPAYLHNRVLPFDSKILTYICGLVALNIPYELP-IYGEDII 125
 CC 117 QYSRHLGELPHIPAIANECYRCMLKRHNDICILKESGAKTESTLILKFLSVISQ 176
 CC 126 AVSGQMGMDPHIFVAEAYKQWABDERNSIIVSGSGAKTVSAKARVATVSG 185
 CC 177 QSLSLKSKTICVERAILSESPIMEAGNATVYNNSSRFGKTVOLNICKGNIOGR 236
 CC 186 SASSEANVEEK-----VLASNPIMESIGNAKTTNDNSRFGKYLEIGFDRYRILGAN 238
 CC 237 IYDCILSSGNRVVRNPGERNHYFALLAGLEHRESEFYLTSTENHYLNQSGVEDK 296
 CC 239 KMTYLL-EKSRVVFQAEERNHYIFQLCASAKLPEFKRLGNADSFYTTQSGSPME 297
 CC 297 TISDSEFEVITAMVQFSKEEVREVSRLAGILHGNIFITAGQVSP---KTAL 353
 CC 298 GVDDAKEMANTQACTLIGISSEYQMGIFRIAGILHGNVGFARSDSCITPRNHEP 357
 CC 354 GSABELGLDPTQLDALTORSMFLRGEILTPLVNQQAVDSRDSLAAALYACCEBWIK 413


```

Db 358 IIFCDIMGVDEEMCHWLCHRKLATATETTYIKPISKLOATNARDALAKHIVAKLFNWIWG 417
Qy 414 KINSRI-KGNEDEFKSIGLIDIFGFENFENVHFEQFINANERKLOEFNKHIFSEOLEY 472
Db 418 HYNOLAHSAVOKHSPIGLVDIYGFETFEINSEFOFINANERKLOQOFMMHVFKEQEBY 477
Qy 473 SREGLWEDIDWIDNGECLDLI EKKLGLLALINSESHFOCATDSTLLEKXHSQHANN-HF 531
Db 478 MKEQIPWTLIDFYDNQPCINLIESKGLIDLDEBCKMFKGTDTMAQGLYVTHLKNAL 537
Qy 532 YKPRVAVNPGVKNYAGVOYDVNGILAKNDTPRDDLLNLRESRPFIYDLFEH--- 588
Db 538 FEKPRMSNRAFLIKHIFADKVEYOCSEGFLEKNKDYTEEQI KYLKSKSKFKMLBQDDBK 597
Qy 589 -----VSSRNQDTLKCSGKRRFPYVSQFKDLSHSLMATS 626
Db 598 AISPTBATSNGTPLTRVPVXPKGRPGT-----AKEHKVYGLQFR-NSLHLMETLN 651
Qy 627 SSNPFVRCIKENMKMPDQFQAVVLANQARYSGMLTYVIRKAGYAVRPPQDFYKRYK 686
Db 652 ATPHYVRCIKENDFPFTFDBKRAVQQLRAQGVLETIRHISAGFPSRWYQEFESRYR 711
Qy 687 VLMKRLALPVDVGRKCTSLDLVDASNSWOLGKTVFLPESLEQLEKREBEVSHAM 746
Db 712 VLMKQDVLGDRKQTCOVNLEKILDKOYOFKTKIFRAGQVATLEKLRADKLRACI 771
Qy 747 VIRAHLGFLARKO-----YRKVL---YCVVLIQKNVAFILR 781
Db 772 RIQKTRIGLRLKRYLCMQRAAITVORVRYGVOARCYAKPLRRTKAATTQKXMYRYVR 831
Qy 782 RRFLLHKKAAIVFOKOLRGQIARVYROLLAE-----KREOE--- 819
Db 832 RYKTKRRAATVYLGSLRYGLYLRNRYKTLREHKAVTIQKAVRGMLARTHYRTKAILY 891
Qy 820 ----KKQEEBEKKKKREERERER-----ERREHLLAQQEE----- 853
Db 892 LOCCFRMMAKRELKCLKIARSVERKYKLIHGMEXKIMQLOKXDEQKDYKCLMEKLT 951
Qy 854 -----ETRK-OOELEALOKSOKAEELT--RELEKOKNNQVEILREKEIEDLOM 902
Db 952 NIEGYVNSTEKLKNDVERLOJSEBAKATGRTVSLQ-----EELAKLRDLETRSE 1005
Qy 903 KEQOELSLTEASLOKQERRQDELRLREBE--ACRAQOELSLNDEIDECEVNIERSL 960
Db 1006 KKS-----IEBRADKYKQETEQVLSNLKXENTLLKQEKTLHLMWEQKEMTEMERLT 1060
Qy 961 SGGSFESSELAESACEKPNF-----NSQPYPEEVEVDEGFADDADFOSPNSEHGHS 1015
Db 1061 ---VETTKOLELDLNDERLRYONLNEFSR--LEERYDD--LKEBMTLMLNVKPP--GH- 1110
Qy 1016 DORTSGIRTSDDS-----SREDPYMNDTVPTSPASDSTVJLAPSVODSGSLNSSGSEST 1071
Db 1111 -KRTDSTHSSNSEYTFSESEPAETEDIAPTBEPTEKKYPLMMSLFLKLOKRVTELQOK 1169
Qy 1072 YCMPQAGDLPSPDGDYDODDYEDGATITSGSVTFSSNSYSGSOWSPDYRCVGTYSNGS 1131
Db 1170 QLM-----QDELD---RKBEQVLRSGKXGGE-RPQIRGHELGESJLK 1207
Qy 1132 AYRFSEGAQSSFEDE-----EDPDSRPTDDELSTYRD 1166
Db 1208 ROELESSENKCLKNELEMLKALSEKAPRVNAPGAPAYRVLMEQLTA---VSEELDVREX 1264
Qy 1167 SYVSCYTLFPHSLPYMKGLNMSW-----KRWCLXDEETLMPFSKQKALKQGLHKK 1221
Db 1265 EV-----LILRSQVLSQKKAIPKODKNTMTDSTLL-LEDVKKMKDKGFIADA 1311
Qy 1222 GGSSTLSRRNWKRFVLRQSLMYFENDSE---EKLKATVEVTRAKIINTTKEKGI 1278
Db 1312 YIGLKEFN-----LLESQLOSQKSHENEMALAN3--EIGSLKE-----EN-- 1351
Qy 1279 DIIIMDRTHLIAES---PEDASQMFVSUQVHASTDOEIQEMHDEQANPON----- 1327
Db 1352 -----NRQOOLLAONQLPPEARIEASLOHEITRLTNEMLDLM--EQLEKQKTVYKCLK 1404

```

```

Qy 1328 -----AVGLTDVGLIDVCSAS---DEPDPSNFVITANKRVLHCNADTEEMHWTTL 1377
Db 1405 QUKVFAKKIGLEFVQGMENISRGQIIDEPINP-----VNIPKKGDFQGM 1449
Qy 1378 LORSKGDTRVBQGEFIVRGMLHKEVNSPKMSSLKTKRWPFVLTNS----- 1424
Db 1450 LEYKREDEQ-----KLVKX---LILELKPGRVAVNLISGLPAILPMCVRH 1492
Qy 1425 LDYKSSSKNMLKGLTVLNSLCSVPPDEKIFKEGTWNVTVYGRKKCYRLYTK-----L 1480
Db 1493 ADYLDQDK-VRSLLTSTINSIKVYKRGODFETVSWLSNTCRFLHCLQKQYSGEBGM 1551
Qy 1481 LNEATRWSSVIONVDTKAPIDTPQOLIODIKENCLN-----SDVEQIYKRN 1529
Db 1552 KHNISR-----QN-----EHCILNFDLAERYQVLSDLAIGIYGO- 1585
Qy 1530 PILRYTHPLHSPPLP--LPYGD-I-NLMLDKGY---TLQDEAIKIFNS-LOQLBSM 1581
Db 1586 -LVRYLENILPMTIVSGMLHEHTIQGVGVKPTGLRKRTSSIADEGYVTLDSILRQLNSF 1644
Qy 1582 SDPIPIIGIOTGHDLAPLDELQCOLIKQNKVPHGVSQNLXSWQILTCLSCTPLPS 1641
Db 1645 HS-----VWCQGMGP---ELIKQVKKQWPIYVGAITLNNL---LLRDMQSW--S 1687
Qy 1642 RGLKYLKFLKRIEOPFTE-MEKVALFTYESL-----KTKCREFPVSRDEI 1690
Db 1688 KCM-QIRVNSQLEWMLRDKLMSGAKTELEPILOAQLQVKKTD-----DDA 1737
Qy 1691 EALIRQEWSTVYCHGGGCKI--TINSHTTAGVEKL-----IRGLAME----- 1735
Db 1738 EATCS-----MCNALTTAQIVKLVNLYPVPNEFEERVSVPFRTIQVRLDRKDSPO 1789
Qy 1736 ---DSRNMFL-FEYN 1747
Db 1790 LMDAKHIFPVTFPPN 1805

RESULT 15
MYS2_YEAST
ID MYS2_YEAST STANDARD; PRT; 1574 AA.
AC P19524;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Myosin-2 isoform.
GN MYO2 OR CDC66 OR YOR326W OR O6167.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_Taxid=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GRF8;
RX MEDLINE=91201404; PubMed=2016335;
RA Johnston G.C., Pendergast J.A., Singer R.A.;
RT "The Saccharomyces cerevisiae MYO2 gene encodes an essential myosin
for vectorial transport of vesicles."
RL J. Cell Biol. 113:539-551 (1991).
RN [2]
RP SEQUENCE OF 1-748 FROM N.A.
RC STRAIN=628BC / FY1679;
RX MEDLINE=97051589; PubMed=8896266;
RA Pearson B.M., Hernandez Y., Payne J., Wolf S.S., Kalogeropoulos A.,
Schweizer M.;
RT "Sequencing of a 35.71 kb DNA segment on the right arm of yeast
chromosome XV reveals regions of similarity to chromosomes I and
XIII."
RL Yeast 12:1021-1031 (1996).
RN [3]
RP SEQUENCE OF 677-1574 FROM N.A.
RX MEDLINE=97051586; PubMed=8896263;
RA Parle-Mcdermott A.G., Hand N.J., Goulding S.G., Wolfe K.H.;

```

RT "Sequence of 29 kb around the PDR10 locus on the right arm of
RT Saccharomyces cerevisiae chromosome XV; similarity to part of
RT chromosome I."
RL Yeast 12:999-1004(1996).
CC -1- FUNCTION: ESSENTIAL MYOSIN FOR VECTORIAL TRANSPORT OF VESICLES.
CC PLAYS A DETERMINANT ROLE IN THE SPATIAL REGULATION OF BUD
CC FORMATION.
CC -1- SUBUNIT: MAY SELF-ASSOCIATE. INTERACTS WITH MLC1.
CC -1- MISCELLANEOUS: CA(2+) MAY PLAY A ROLE IN MYO2 ACTIVITY.
CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M35532; AAA34810.1; -
DR EMBL: X30565; CA62184.1; -
DR EMBL: 249821; CA68973.1; -
DR EMBL: 275234; CA93646.1; -
DR EMBL: 275235; CA93648.1; -
DR PIR: A38454; A38454.
DR HSSP: P10587; 1BR2.
DR SGD: S0005853; MYO2.
DR InterPro: IPR002710; DIL.
DR InterPro: IPR000048; IQ_region.
DR InterPro: IPR004009; Myosin N.
DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF00612; IQ; 6.
DR Pfam: PF01843; DIL; 1.
DR Pfam: PF02736; Myosin N; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR ProDom: PD003376; DIL; 1.
DR SMART: SM00015; IQ; 2.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
DR Myosin; Coiled coil; ATP-binding; Actin-binding; Alkylation;
KW Multigene family; Phosphorylation; Calmodulin-binding.
FT DOMAIN 1 781 MYOSIN HEAD-LIKE.
FT 832 861 IQ.
FT DOMAIN 926 981 COILED COIL.
FT DOMAIN 982 1009 HINGE.
FT DOMAIN 1010 1086 COILED COIL.
FT DOMAIN 1087 1574 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
FT DOMAIN 443 523 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 750 925 ARG/LYS-RICH (BASIC).
FT DOMAIN 791 805 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 814 828 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 839 853 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 862 876 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 887 902 CALMODULIN-BINDING (BY SIMILARITY).
FT DOMAIN 910 924 CALMODULIN-BINDING (BY SIMILARITY).
FT NP_BIND 164 171 ATP (BY SIMILARITY).
FT MOD_RES 692 692 ALKYLATION (BY SIMILARITY).
FT MOD_RES 702 702 ALKYLATION (BY SIMILARITY).
FT MOD_RES 1097 1097 PHOSPHORYLATION (BY CDC2) (POTENTIAL).
SO SEQUENCE 1574 AA; 180680 MW; 1F7E2887C1E59D54 CRC64;

Query Match 13.4%; Score 1442; DB 1; Length 1574;
Best Local Similarity 32.5%; Pred. No. 9, 2e-59;
Matches 311; Conservative 228; Mismatches 413; Indels 130; Gaps 29;

QY 55 HPNKGVDMDMLTLLHGSGIMYNLFQYKRNQITTYIGSILASVNPQPIAGLYEPAT 114
DB 65 NPPILEATEDLSTLYLNEPAVLAHKQYSQILNITVYSGIVLIATNPFDVQDLTYTDM 124

QY 115 MEQYSRRLGELPRPIFAIANECYRCWKMDNOICILKSGESGAKTESTKILK-FLSV 173
DB 125 IQAVGKRGELPRPLFAIAEAYRLMKDQKNGITVSGESGAKTYSAKITMYFASV 184
QY 174 IQQSLSLSKKTECVERPAIIESSPIAEFQNAATVYNNNSRGGKYVQNLICQKQINQ 233
DB 185 EENENATVOHQVEMSETEQKILATNPIMEAFQNAATTRDINSRGGKYLEIFDQDTSIT 244
QY 234 GGRIVDCILSSQNRVVRQNGPERNYHIFALLAGLEHERREFEVYSTPENHYHYNQSGCV 293
DB 245 GKRITTYLL-ERSLVYQPPRIERNHYIFQLAGAPACQKEELHLDASDYTYMNGGDT 303
QY 294 EKRISDSESPREVITANDVWQFSKEVEVSRLLAGILHGNIEF-ITAGAQVSP-KT 351
DB 304 KINGIDAEKYEKITVADLTALVIGITKETOHQIKLIALHLHICNIEIKKTRNDASISADER 363
QY 352 ALGRSAELLGLDPTQUTDALTRSMFLGEEILITPLANQOAVSDSLMALIACCFEV 411
DB 364 NKLACELGLIDAVNPAKVKVTKQITRSEKIVSNLNSQALVAVDSVAKFTYSALFDWL 423
QY 412 IKKINSRIKG---NDEPKS-IGILIDIFGFENFEVNHFEQFINVANEKLOEYFNKHIFSL 467
DB 424 VENINTVLCNPAVNDQISSFVIGVDIYGFHEFEKKSPEQFCINANEKLOEYFNQHVFL 483
QY 468 EOLEYSRGLVWEDIDWIDNGECLDIEKGLALLINEESHFPQATDSTLLEKLH---S 524
DB 484 EOEVEYKEIEWSFIEFNDQPCDILNKGITLSLDEBSRLPGSDPSWTQKLYQTL 543
QY 525 QHANNHFYKPRVAVANNQGVKAYAGEVQYDVRGILEKXNDTRDILLNLSRDFIYD 584
DB 544 KSPTKVPSKPRFGQTKFVSHYADVAVDGFIEKNDVYSDHLEVLKASTETTLN 603
QY 585 LFEHY--SSRNNQDTLK-----CGSK-----HRRPTVSSQFYDSHSMATLSSN 629
DB 604 ILEGLEKAAKKLEBAKKLEBQAGSKKQPRITVRNRPFLSGMFK-QSILEMNTINSN 662
QY 630 PEFVACIKNNQKMDQFDQAVVNLQRLYSGLMLETVRIRKAGVAVRPFQDYKQYKVL 689
DB 663 VHYICIKPNADKEMQFQNLNVLQLRACGVLETIRISCAQFPMRWTFEEVLRYYLI 722
QY 690 -----RNLALPEDYRGKCTSLDLYDASNSEMQLGKTVLRBSLEKLRRE 739
DB 723 PHEQWDLIFKKKETEEDDISVKNMILDAVYKDKSKYQDGNKIFPKGMLAYLELRSN 782
QY 740 BYSHAMVIRAVLFGRLAKOYKRVLYCVIIOKYRPAFLRRFLHLK--AAIVFQ 797
DB 783 KQNSIWMIOKKIRAKYTRKQYLQISQAIKYLQNNIKGFIIRQVNDENKVCATLLQA 842
QY 798 LRG-QIARRVYQLAEKREBEKKKQEBEKKREBE-----EREREREAEELR 848
DB 843 YRGHSIRANVFVLRTITNLQKKIRKELKQRLQKQHEVNAVLIQSKVRFEPSPRFLR 902
QY 849 AQOE-----EETKQOEALQKS-----QKAEELTREL-EKQEK 884
DB 903 TKKDVVVOSSLRRRAQRKQLKQADAKSVNHLKESYKLENKYIELQNLASKYKQEK 962
QY 885 -----QVEELRLKEIE-----DLORKKE-----QOELSTEASQ 916
DB 963 EMTREIKELQVVEESAKIQEITLNMKKENHLIDINQSKMDELQKITENNIOSTEQTLK 1022
QY 917 KIQ-----ERRQOELARLEE-EACRAOEFLESLND-----EIDECVNRIR 958
DB 1023 DAQLELDVWQKHDELKESKKQLEBELQTKTLVVEYQTLNDQNEVSKLEELARLT 1082
QY 959 SLGSGSEFSSELAESAACEKPN--FNFSQPIPEEVEDEGFADDAFADFSNPSSE-HGH 1014
DB 1083 AMSLGTIVTTSVLPQPLPLDWMGGASNFNNMWE--NSDSLSPNDLNLKSRSTPSSGNH 1139
QY 1015 SD-----QRTSGIRTDSDSEEDPVYNDPVVPSPADSYVLLAPSVQDSGLHNSGGS 1070
DB 1140 IDSLSVDRENGVNAQINEELYRLLEDTIELNQEITEG-LLGFEVPPDGAVALQSKRDV 1198
QY 1071 TY 1072

Db 1199 VY 1200

Search completed: July 14, 2003, 18:16:57
Job time : 43.3952 sec

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:07:56 ; Search time 95.4296 Seconds

(without alignments)
4441.379 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730

Sequence: 1 PCLQCTRWLRNGQHFPST.....MIVKKRYSTTRSSSQGSSR 2057

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaea:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10011	93.3	2062	11	Q91JY5 mus musculus
2	4744	44.2	924	11	Q8R3S0 mus musculus
3	3294	30.7	712	4	Q94893 homo sapien
4	1921	17.9	2357	5	Q9U0M8 Q9U0M8 dicystosell
5	1843.5	17.2	2167	5	Q9VJZ6 Q9VJZ6 drosophila
6	1834.5	17.1	2179	13	Q9DGG8 Q9DGG8 brachydanio
7	1831.5	17.1	2179	13	Q9DGG9 Q9DGG9 brachydanio
8	1716	16.0	2098	5	P91443 P91443 caenorhabdi
9	1705	15.9	2129	5	Q9VJZ3 Q9VJZ3 drosophila
10	1697	15.8	2121	5	Q9NH54 Q9NH54 drosophila
11	1635	15.2	2113	11	Q99M26 Q99M26 mus musculus
12	1512.5	14.1	1529	10	Q9XE14 Q9XE14 zea mays (m
13	1495	13.9	2424	5	Q9VZ48 Q9VZ48 drosophila
14	1492	13.9	2548	4	Q9UNJ2 Q9UNJ2 homo sapien
15	1487.5	13.9	1611	10	Q8RYE8 Q8RYE8 arbidopsis
16	1487	13.9	2626	11	Q9Z1N3 Q9Z1N3 rattus norv

17	1485	13.8	1260	10	Q24516 Q24516 helianthus
18	1483	13.8	1528	10	Q24518 Q24518 helianthus
19	1480.5	13.8	1533	10	Q8S0C8 Q8S0C8 cryza savi
20	1466	13.7	1511	10	Q9M5A6 Q9M5A6 valianeria
21	1463	13.6	1547	10	Q8W312 Q8W312 cryza savi
22	1458.5	13.6	1477	10	Q64491 Q64491 arbidopsis
23	1458	13.6	1520	10	Q39160 Q39160 arbidopsis
24	1450.5	13.5	1556	10	Q9ZVN3 Q9ZVN3 arbidopsis
25	1449.5	13.5	2148	5	Q9ATB7 Q9ATB7 petroselinu
26	1441.5	13.4	1515	5	Q9BDJ3 Q9BDJ3 physarum po
27	1440	13.4	1505	10	Q9LKB9 Q9LKB9 arbidopsis
28	1437	13.4	1824	5	Q9NBH3 Q9NBH3 strongyloce
29	1435.5	13.4	1599	10	Q9LNP7 Q9LNP7 arbidopsis
30	1435	13.4	1515	10	Q39158 Q39158 arbidopsis
31	1434.5	13.4	1446	10	Q9SVT9 Q9SVT9 arbidopsis
32	1431	13.3	2651	10	Q9FR85 Q9FR85 arbidopsis
33	1430.5	13.3	1490	10	Q9SK80 Q9SK80 arbidopsis
34	1430	13.3	2167	10	Q9SUL1 Q9SUL1 chara coral
35	1430	13.3	2182	10	Q9LW97 Q9LW97 chara coral
36	1424.5	13.3	1242	10	Q9M2K0 Q9M2K0 arbidopsis
37	1423	13.3	1374	10	Q9SMY9 Q9SMY9 arbidopsis
38	1399.5	13.0	1502	10	Q24517 Q24517 helianthus
39	1399.5	13.0	1736	10	Q23025 Q23025 arbidopsis
40	1399	13.0	1516	3	Q94477 Q94477 echinzoacch
41	1394	13.0	1502	10	Q9SK73 Q9SK73 arbidopsis
42	1392	13.0	1285	4	Q9BZ27 Q9BZ27 homo sapien
43	1391	13.0	1285	4	Q9UEG2 Q9UEG2 homo sapien
44	1386.5	12.9	1304	13	Q9PWF6 Q9PWF6 morone saxa
45	1379	12.9	1471	3	Q74805 Q74805 schizosacch

ALIGNMENTS

RESULT 1	Q9UJY5	PRELIMINARY:	PRT:	2062 AA.
AC	Q9UJY5;	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Myosin X.			
GN	MYO10 OR MYO 10.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=BALB/C;			
RC	MEDLINE=20261293; PubMed=10799329;			
RA	Yonezawa S., Kimura A., Koshida S., Masaki S., Ono T., Hanai A.,			
RA	Somta S., Kageyama T., Takahashi T., Moriyama A.;			
RT	"Mouse myosin X: Molecular architecture and tissue expression as			
RT	revealed by northern blot and in situ hybridization analyses.";			
RL	Biochem. Biophys. Res. Commun. 271:526-533(2000).			
DR	EMBL; AJ249706; CAB56466.2; -			
DR	HSSP; P08789; 1AMD.			
DR	MGD; MGI:107716; Myo10.			
DR	InterPro; IPR000299; Band_4.1.			
DR	InterPro; IPR000048; IQ_region.			
DR	InterPro; IPR001609; myosin_head.			
DR	InterPro; IPR001649; PH.			
DR	Pfam; PF00612; IQ_3.			
DR	Pfam; PF00603; myosin_head; 1.			
DR	Pfam; PF00784; MYTH4; 1.			
DR	Pfam; PF00169; PH; 3.			
DR	PRINTS; PR00193; MYOSINHEAVY.			
DR	PRODom; PD000355; myosin_head; 1.			
DR	SMART; SM00295; B41; 1.			
DR	SMART; SM00015; IQ; 3.			
DR	SMART; SM00242; MYSC; 1.			

```
DR SMART; SM00139; MYTH4; 1.  
DR SMART; SM00233; PH; 2.  
DR PROSITE; PS50057; BAND_41_3; 1.  
DR PROSITE; PSS0096; IQ; 1.  
DR PROSITE; PSS0003; PH DOMAIN; 2.  
SQ SEQUENCE 262 AA; 23746 MW;
```

Query Match	93.3%	Score 10011	DB 11	Length 2062
Best Local Similarly	93.3%	Pred. No. 0		
Matches 125; Conservative	64	Mismatches	64	Indels 10; Gaps 6

Qy	1	CLDQTRWMLBNGOHPSSTVNSCAEGIVPFTDYGQVFTYQOSITTHOKRATAMPTEE	60
Db	4	FPBEGARWMLBNGOHPPSPVNSCAGVVPFTDYGQVFTYQOSITTHOKRATAMPTEE	63
Qy	61	GVDDMASLTELHGSGIMYNLFQRYKRNQIMWTYISGILASVNPQPIAGLYEPATMEQYSR	120
Db	64	GVDDMASLTELHGVAIMYNLFFQRYKRNQIMWTYISGILASVNPQPIAGLYEPATMEQYSR	123
Qy	121	RHLGELPHFIRAIANEYCRLMKRDNOCILIKSGSGAKTESTKULIKPLSVIQOSLE	180
Db	124	CHGELPHFIRAIANEYCRLMKRDNOCVILSGSGAKTESTKULIKPLSVIQOTLD	183
Qy	181	LSLAKTSCYBRALIESSPIMEAFNATVYNNNSRRGKFPQULNICQGNIGGRIVDC	240
Db	184	LGLQKTSVVEQALIQSSPIHEAFNATVYNNNSRRGKFPQULNICQGNIGGRIVDY	243
Qy	241	ILSSONRVVRONGPGRNHYIFAYALLAGLEHEEREEFPYLSTPENHYLNQSGCEDKTTSD	300
Db	244	LL-EKNRVVRRIIPGRNHYIFAYALLAGLQDEEREEFPYLSTPENHYLNQSGCEDKTTSD	302
Qy	301	QESPREVITANDWQFSKEEYREVSRLLAGILHLNIEFITPAGQVQSFKTALGSAELL	360
Db	303	QESPRQVITAEWQFSKEEYREVIRLLAGILHLNIEFITPAGQVQIPKTKALGSAIDL	362
Qy	361	GLDPTQLDALTORSMFLRGEIILPLNVQQAVIDSRDLSLAMLVYACCEFWIKKINSRIK	420
Db	363	GLDPTQLDALTORSMILRGEIILPLSVQQAVIDSRDLSLAMLVYARCEFWIKKINSRIK	422
Qy	421	GNEDFKSGIILIDFGFENFVNHFFQFININANEKULOEFYFNKHTISLEOLEYSGELWYE	480
Db	423	GKDFPKSGIILIDFGFENFVNHFFQFININANEKULOEFYFNKHTISLEOLEYSGELWYE	482
Qy	481	DIDIMDNECJDLEKJGLLALINESHPPOATSTLTLEKHSOAHNNHFVYKRVAVN	540
Db	483	DIDIMDNECJDLEKJRLALINESHPPOATSTLTLEKHSOAHNNHFVYKRVAVN	542
Qy	541	NFGVGHVAGEVOYDVRGILEKNRDITFRDULNLRESRDFIYDLFEHVSNNODTLKC	600
Db	543	NFGVGHVAGEVOYDVRGILEKNRDITFRDULNLRESRDFIYDLFEHVSNNODTLKC	602
Qy	601	GSKRRRPVSSQFKUDLSHLSMATLSSSNPFVRCIKRMQMPQPODAVVLNOLARISG	660
Db	603	GSKRRRPVSSQFK-DLSHLSMATLSSSNPFVRCIKRMQMPQPODAVVLNOLARISG	661
Qy	661	MLETVIRIKAGAVARPPODFYKRYKVMRLNALPEDVRGKTSLLQDYDASNSMOQK	720
Db	662	MLETVIRIKAGAVARPPDFYKRYKVMRLNALPDPDRGKTVLLQDYDASNSMOQK	721
Qy	721	TKVPLRESLEOKLEKXREEBVSHAAMVIRAHVGLGFIAKQYRKVLYCVIILQKNYRAEL	780
Db	722	TKVPLRESLEOKLEKXREEBIDRAMVIRAHVGLGFIAKQYRKVLCGVTTIQKNYRAFLA	781
Qy	781	RRRFLHLKKAALIVQOKLRGQIARVYVQULLAEKREOEKKQOESEBKKKREEREBER	840
Db	782	RKKFLHLKKAALIVQOKLRGQIARVYVQULLAEKRELEBKKKREEREBER	839
Qy	841	ERRBAE-LRAQOEETTRQOELALQOKSEKELFRELLEKOKENQOVEIIRLEGEIDL	899
Db	840	AQREBDILRAQOEAETTRQOELALQOKSGRAADLTRELEKQRENQOVEIIRLEGEIDL	899
Qy	900	QRMKQOELSTEASLOQLQERRDQELRLEBEACRAAOEPLSINPDEIDECVNIERS	959

Db	900	QRMKERGELSTIEASLOKLOOLRBEELRLRLEDEACRAAOFELBSEANDEIDECVARNIERS	955
Qy	960	LSGSEBFS---SELSASACEKCPNFNSQPyP-EZEVDSGFADDDAFKOSPNPSEHG	1014
Db	960	LSVSEISIGEELSEIASESAQEKSPFNFSQPyPEEVEDEGFADDDAFKOSPNPSEHG	1019
Qy	1015	SDQRTSGIRTSDDSEEDPYNMPTVPPSPADSTVLLAPVQDSGLSHNSQSESTYCM	1074
Db	1020	SDQRTSGIRTSDEESSEEDPYNMPTVPPSPADSTVLLASMQDSALSHNSQSESTYCM	1079
Qy	1075	PONAGDLPSPGDYVDODDYEDCAITSGSSVTESNSYGQMSPDYCSVGTYNSSGAYR	1134
Db	1080	PONEGDLPSPGDYVDODDYEDCAITSGSSVTESNSYGQMSPDYXSVGTYNSSGAYR	1139
Qy	1135	FSSSEGAQSSFFDSSEDFDSRFDTDDELSTYRDSVYSCVTLPyFHSFLYMKGLMNSWKRR	1194
Db	1140	FSSSEGAQSSFFDSSEDFDSRFDTDDELSTYRDSVYSCVTLPyFHSFLYMKGLMNSWKRR	1199
Qy	1195	WCULKDEFTFLMFRKQEAHLKQGMHLKKGGSSTLSRNMWKRMVLLNOSKLTMYENDSEE	1254
Db	1200	WCVLKDEFTFLMFRKQEAHLKQGMHLKKGGSSTLSRNMWKRMVLLNOSKLTMYENDSEE	1259
Qy	1255	KLKGTVEVRTKEIJIIDNTKENGDIIMADSTPHLASEPDAQOMFVSLSQVASTDQE	1314
Db	1260	KLKGTVEVRTKEIJIIDNTSKENGDIIMADSTPHLASEPDAQOMFVSLSQVASTDQE	1319
Qy	1315	IOENHDEQANPQNAVGLTDVGLIDSVCAQSDSPDRPNSFVIITANRVLHCNADTEEMHHW	1374
Db	1320	IRENHDEQANPQNAVGLTDVGLIDSVCAQSDSPDRPNSFVIITANRVLHCNAYTDEEMHHW	1379
Qy	1375	ITLLQRSKGDTRVSGOEFYVNGMYLHKEVXNSPKXNSLKLKRMVVLTHNSLDYTKSSEKN	1434
Db	1380	ITLLQRSKGDTRVSGOEFYVNGMYLHKEVXNSPKXNSLKLKRMVVLTHNSLDYTKSSEKN	1439
Qy	1435	ALKXGLTVLNSLGSVNPDEKIKFETGYNMNTVYGRNHCVRLYLKLNEATRMSASSYONV	1494
Db	1440	ALKXGLTVLNSLGSVNPDEKIKFETGYMNTVYGRNHCVRLYLKLNEATRMSASSYONV	1499
Qy	1495	TDTRAPIDTPTQOOLIODIKENCCLNSDVAVEQIYKKNPILRYTHHPLSHPLPLPYGDINLN	1554
Db	1500	TDTRAPIDTPTQOOLIODIKENCCLNSDVAVEQIYKKNPILRYTHHPLSHPLPLPYGDINLN	1559
Qy	1555	LLKXGTTTTLQDEAKIKIPNSIQOLESMSDPPIIOGILQTHDRLPLRDELYCQLIKOTN	1614
Db	1560	LLKXGTTTTLQDEAKIKIPNSIQOLESMSDPPIIOGILQTHHRLPLRDELYCQLIKOTN	1619
Qy	1615	KVPHRGSVGNLYSNOILITCLSCFPLPSNGILAKYLKPHLKRREOPGPTBMEKVALPTYES	1674
Db	1620	KVPHRGSVGNLYSNOILITCLSCFPLPSNGILAKYLKPHLKRREOPGPTBMEKVALFYES	1679
Qy	1675	LKKTCKREFFVSPREIEALHROEMTSIVYCHGGGSKCITINSHTTAGEVVEKLIIRGLAM	1734
Db	1680	LKKTCKREFFVSPREIEALHROEMTSIVYCHGGGSKCITINSHTTAGEVVEKLIIRGLAM	1739
Qy	1735	EDSRNMFALFEYNGHVDAIESRTVADVLAKEFKLAATSEVGLPMKFYFKLYCFLDTD	1794
Db	1740	EDSRNMFALFEYNGQVKAIESRTVADVLAKEFKLAATSEAGDAPMKFYFKLYCFLDTD	1799
Qy	1795	NVPMDSVBFAPMEQOAHAVHGHHPAPREMLQVLAALRLQYLOGDVTLMAHAPRELEVY	1854
Db	1800	SMPPDQGEAFAPMEQOAHAVHGHHPAPREMLQVLAALRLQYLOGDVTLPHSIPRELEVY	1859
Qy	1855	SLQRLKARISGSTYKTPCERLSEKRRAPSPLEGTJARSFRTGSVVRQKVEEQMDMIMIKE	1914
Db	1860	SVQRLRARISSSTYKTPFYERLEKRRKSPFLEGTJARSFRTGSVVRQKAAEEQMDMIMIKE	1919
Qy	1915	EVSASARASIIDKMKRFQGMNOBOAMAKYMLIKEMPGYSTLFDVECKEGGFQDELMLGV	1974
Db	1920	EVCASARASIIDKMKKLOGMQBOAMAKYMLIKEMPGYSTLFDVECKEGGFQDELMLGV	1979
Qy	1975	SADAVSYTKGEGAPLEVFOYEHLTSGAPLANTYKIVDERELLFETSEVVDYAKLMA	2034
Db	1980	SABAVSYTKGEGAPLEVFOYEHLTSGAPLANTYKIVDERELLFETSEVVDYAKLMA	2039

QY 2035 YISMIYKRYSTTRSSASQSSSR 2057
 DB 2040 YISMIYKRYSTTRSSASQSSSR 2062

RESULT 2

QY 098350 PRELIMINARY; PRT; 924 AA.
 AC 098350;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DE Similar to myosin X (Fragment).
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxId=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Strauberg R.;
 DR EMBL; BC024692; AAH24692.1; -.
 FT NON_TER
 SQ SEQUENCE 924 AA; 106409 MW; F4F4E9E4A1C3819B CRC64;

Query Match 44.2%; Score 4744; DB 11; Length 924;
 Best Local Similarity 97.2%; Pred. No. 1.1e-241;
 Matches 888; Conservative 13; Mismatches 13; Indels 0; Gaps 0;

QY 1134 RRSSEGAQSSFEDESDPDSRPDTDELSTYRDSVYSCVTLPYFHSFLYKMGGLMNSMKR 1193
 DB 1 RRSSEGAQSSFEDESDPDSRPDTDELSTYRDSVYSCVTLPYFHSFLYKMGGLMNSMKR 60
 QY 1194 RMCVLTDETFMFRSKQEQALKQGMHLKGGSSSTLSRRNKKRWFLVLRQSKMTYFENDSE 1253
 DB 61 RMCVLTDETFMFRSKQEQALKQGMHLKGGSSSTLSRRNKKRWFLVLRQSKMTYFENDSE 120
 QY 1254 EKLKGTVEYRTAEIIDNTTKENGIDIMADRTFHLIAEPEASQMFVLSOVHASTQO 1313
 DB 121 EKLKGTVEYRTAEIIDNTTKENGIDIMADRTFHLIAEPEASQMFVLSOVHASTQO 180
 QY 1314 EIOEMHDEQANPQNAVGLDVGILDSVCSDSDBRPNFVITITANVYLHCNADTPEEMH 1373
 DB 181 EIREMHDEQANPQNAVGLDVGILDSVCSDSDBRPNFVITITANVYLHCNADTPEEMH 240
 QY 1374 WITLLQRSKGDTRVEQGEPIVRGMLHKEVKNSPKMSLKKKRPVLTTHNSLDYKSSSEK 1433
 DB 241 WITLLQRSKGDTRVEQGEPIVRGMLHKEVKNSPKMSLKKKRPVLTTHNSLDYKSSSEK 300
 QY 1434 NLLKGLTIVLNSLCSVVPPEDEKIFKGTGVNNTVYGRKHCYRLYTLLNEATRWSSVION 1493
 DB 301 NLLKGLTIVLNSLCSVVPPEDEKIFKGTGVNNTVYGRKHCYRLYTLLNEATRWSSVION 360
 QY 1494 VTDTKAPIDPTQOOLQIDIKENCINSDVYEQIYKRNPILEYTHHPLHSPILPLPYGDINL 1553
 DB 361 VTDTKAPIDPTQOOLQIDIKENCINSDVYEQIYKRNPILEYTHHPLHSPILPLPYGDINL 420
 QY 1554 NLLKDGYYTTLQDEBAIKIFNSLQOLESMSDPPIPIQGIILCTGHDLPRLDELVCOLIKOT 1613
 DB 421 NLLKDGYYTTLQDEBAIKIFNSLQOLESMSDPPIPIQGIILCTGHDLPRLDELVCOLIKOT 480
 QY 1614 NKVPHGSVGNLYSMOILCTCLPISRGILKYLPFKHRIEOPGTMEYALPTYE 1673
 DB 481 NKVPHGSVGNLYSMOILCTCLPISRGILKYLPFKHRIEOPGTMEYALPTYE 540
 QY 1674 SLKTKICREFVPSRDEIEALIHROEMTSTYCHGSGCKITINSHTAGSEVEKLIRGLA 1733
 DB 541 SLKTKICREFVPSRDEIEALIHROEMTSTYCHGSGCKITINSHTAGSEVEKLIRGLA 600
 QY 1734 MEDSRNMFALFEYNGVHDKAIESRTVAVDLAKPEKLAATSEVGDLPWKYFPLCYGFLDT 1793
 DB 601 MEDSRNMFALFEYNGVHDKAIESRTVAVDLAKPEKLAATSEVGDLPWKYFPLCYGFLDT 660

QY 1794 DNVPSDVEFAFMFEQAEAVIHHHPAPEBNLQVLAALRLQYLOGDYTLTAALPPELEV 1853
 DB 661 DSMPSDVEFAFMFEQAEAVIHHHPAPEBNLQVLAALRLQYLOGDYTLPHSTIPPELEV 720

QY 1854 YSLQRLKARISOSTTFTPTPCERLEKRTSPLEGTIRRSFRGVSIVRKAYEEOMLDMNIX 1913
 DB 721 YSVQRLRARISOSTTFTPTPCERLEKRTSPLEGTIRRSFRGVSIVRKAYEEOMLDMNIX 780

QY 1914 BEVSSARASIIDKMRKFGQMNQEQAMAKYMLIKEMPGVSTLPVECKEGGFPOELNIG 1973
 DB 781 BEVSSARASIIDKMRKFGQMNQEQAMAKYMLIKEMPGVSTLPVECKEGGFPOELNIG 840

QY 1974 VSADAVSYKRGGRPLFVFOYEHILSPGAPLANTYKIIVDERELLFTSSEVVDVAKLMK 2033
 DB 841 VSADAVSYKRGGRPLFVFOYEHILSPGAPLANTYKIIVDERELLFTSSEVVDVAKLMK 900

QY 2034 AYISMIYKRYSTTRSSASQSSSR 2057
 DB 901 AYISMIYKRYSTTRSSASQSSSR 924

RESULT 3

QY 094893 PRELIMINARY; PRT; 712 AA.
 AC 094893;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE KIAA0799 protein (Fragment).
 GN KIAA0799.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NC NCB1_TaxId=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRN;
 RX MEDLINE=98087487; PubMed=9872452;
 RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
 RA Kotani H., Nomura N., Ohara O.,
 RT "Prediction of the coding sequences of unidentified human genes. XI.
 RT for large proteins in vitro."
 RL DNA Res. 5:277-286 (1998).
 DR EMBL; AB018342; BAA34519.1; -.
 DR InterPro; IPR000299; Band_4.1.
 DR InterPro; IPR000857; MYTH4.
 DR Pfam; PF00784; MYTH4.1.
 DR Pfam; PF00169; PH.1.
 DR SMART; SM00295; B41.1.
 DR SMART; SM00139; MYTH4.1.
 DR SMART; SM00233; PH.1.
 DR PROSITE; PS50057; BAND_41.3; 1.
 DR PROSITE; PS50057; PH_DOMAIN; 2.
 FT NON_TER
 SQ SEQUENCE 712 AA; 82011 MW; CE59A9072350B8F5 CRC64;

Query Match 30.7%; Score 3294; DB 4; Length 712;
 Best Local Similarity 99.4%; Pred. No. 1.9e-165;
 Matches 624; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1337 ISVCSASDSDPDRPNFVITITANRVLHCNADTPEEMHWTLLQRSKGDTRVGEQETVVG 1396
 DB 1 ISVCSASDSDPDRPNFVITITANRVLHCNADTPEEMHWTLLQRSKGDTRVGEQETVVG 60
 QY 1397 WLLHKEVKNPKNSSSLKLLKKRWFLTHNSLDYKSSSEKVALKLGTLVNLSCSVPPDEKI 1456
 DB 61 WLLHKEVKNPKNSSSLKLLKKRWFLTHNSLDYKSSSEKVALKLGTLVNLSCSVPPDEKI 120
 QY 1457 FKETGYNNTVYGRKHCYRLYTLLNEATRWSSVIONVTDTKAPIDPTQOOLQIDIKENC 1516

Db	121	PKRGYNAVTVYGRGHGCVRLVTKLLNLEATRMSAIIQAVTDTIKAPIDPTFOGLIDIKENC	180
Qy	1517	LNSDVVEQIYKRNPIRLRTTHHPPLHSPPLPLRYGGINLNLKDGKTYTIOEALIKFNSLQ	1576
Db	181	LNSDVEEIQKRNPIRLRTTHHPPLPLRPYGINLNLKDGKTYTIOEALIKFNSLQ	240
Qy	1577	QLESMSDPIPIIIOGLTQGHDLPLRDELVCOLIKOTNKVPHPGSVGNLYSMQILTCLSC	1636
Db	241	QLESMSDPIPIIIOGLTQGHDLPLRDELVCOLIKOTNKVPHPGSVGNLYSMQILTCLSC	300
Qy	1637	TFLPBGRILKYLKFKLAKIREQFPGTEMEKXALFTYESLKKTKCREFPVPSDEIEALHR	1696
Db	301	TFLPBGRILKYLKFKLAKIREQFPGTEMEKXALFTYESLKKTKCREFPVPSDEIEALHR	360
Qy	1697	QEMSTVYCHGGGSGSKITINSHTTAGEVVEKLIINGLAMEDSRNMPALPEVNGHDKLIES	1756
Db	361	QEMSTVYCHGGGSGSKITINSHTTAGEVVEKLIINGLAMEDSRNMPALPEVNGHDKLIES	420
Qy	1757	RTVADVLAKEBEKLAATSEVDLPMKFEFKLYCFELDTDNVSKDSVEPAFMPEQAHNAVH	1816
Db	421	RTVADVLAKEBEKLAATSEVDLPMKFEFKLYCFELDTDNVSKDSVEPAFMPEQAHNAVH	480
Qy	1817	GHPHPEENLOVLAALRLQYIOGDYTLHAIIPPLPEEYVSLORLKRISQSTKTTTPCERL	1876
Db	481	GHPHPEENLOVLAALRLQYIOGDYTLHAIIPPLPEEYVSLORLKRISQSTKTTTPCERL	540
Qy	1877	EKKRRTSFEGLTLRSRFRTSVVRQKVEBEOMLDMWIKEEVSSARASIIDKKRKQGNNO	1936
Db	541	EKKRRTSFEGLTLRSRFRTSVVRQKVEBEOMLDMWIKEEVSSARASIIDKKRKQGNNO	600
Qy	1937	QAMAKYMALIKEMPEYGSTLPDVECKEG	1964
Db	601	QAMAKYMALIKEMPEYGSTLPDVECKEG	628
RESULT 4			
Qy	Q09JIM8	PRELIMINARY;	PRT, 2357 AA.
AC	Q09JIM8;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	Class VII unconventional myosin.		
GN	MyoI.		
OS	Dictyostelium discoideum (Slime mold).		
OC	Eukaryota; Mycetozoa; Dictyosteliales; Dictyostelium.		
OX	NCBI_TaxId=44689;		
RN	[1]		
RP	SEQUENCE OF 254-349 FROM N.A.		
RX	MEDLINE=95023928; PubMed=7937787;		
RA	Titus M.A., Kuspa A., Loomis W.F.;		
RT	"Discovery of myosin genes by physical mapping in Dictyostelium.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 91:9446-9450(1994).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20045026; PubMed=10574761;		
RA	Titus M.A.;		
RT	"A class VII unconventional myosin is required for phagocytosis.";		
RL	Curr. Biol. 9:1297-1303(1999).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RA	Titus M.A.;		
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.		
CC	-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.		
DR	EMBL; L35321; AAF06035.1; -.		
DR	HSSP; P08799; IIMD.		
DR	InterPro; IPR000299; Band 4.1.		
DR	InterPro; IPR000048; IQ_region.		
DR	InterPro; IPR001609; myosin_head.		
DR	InterPro; IPR000857; MYTH4.		
DR	InterPro; IPR000159; RA_domain.		
DR	InterPro; IPR001452; SH3.		
DR	Pfam; PF00612; IQ; 3.		

[illegible]

Db 780 SNNALVLRDRNARMLEIÖREKOEERNÖKEERDÖKEDEKEDKEXETADRROLOEÖQKR 839
 Qy 858 OOELEALOKSÖKEAELTRELKEKÖKENÖVEILTLEKEJEDLÖRM-KEÖOEILSTLEA--- 913
 Db 840 BEELBA---KREEBELKLEKEKSÖKLENOI---DELSLERMLKEQÖDKNIELDDF 892
 Qy 914 -SLQ-----KLOERDÖELRLLE-BAÖRAÖEFLSINPEIDEC 952
 Db 893 VNSLEAFPEGGVDDSQPYSPNNHÖYEMSPBALDKISTIDLÖGKÖQTRASTKEFEVDES 952
 Qy 953 VANIERISJSGSEFSEELASÖCEKKNNSÖQPIPEEBVDCGFADDAFÖDSPNSBH 1012
 Db 953 ---KPELPPGIENVLKRAPIKROQASFLPGQPIPDVYSPOYPVDE----- 996
 Qy 1013 GHSÖDRTSGIRTSÖSSEDPYNN---DTVPTSPADSIVLAPSVÖDSSGLH----- 1063
 Db 997 ---ADDDSNNTYINSNGDLPPLTSÖSDPSILPPPS---SSMDPGLPP 1041
 Qy 1064 -NSSSGESTYCPÖNAGDLPSPDGDYDÖDDYEDGA-----ITSGSVTSNSYGSQW 1116
 Db 1042 PPPSSSSGGTSLP-----PWP--VFDFGMIDPLGAPPPPTSDSTSPSATATGN-- 1091
 Qy 1117 SPDYCSVGTVNSGAYRFSSEGAÖSS-----FEDSEEDFSRFD--- 1156
 Db 1092 ---NTPNSS---SASASÖSTNQVNPÖPVSVELPÖILNDEEISLSPFYDAN 1138
 Qy 1157 ---TDELSTYRDSYVSCY-----TLPYFHS----- 1179
 Db 1139 KNPIEBKÖKÖDDIFSYÖKSHIKSLLVHSDBÖQKVAVEITFSKVLYHNSPLVSKÖP 1198
 Qy 1180 ---FLYMKGGLMNSMKRMÖCVLKDETF-----LW----- 1205
 Db 1199 ADPYSPVKEFILTKGLAIBS-----LRDEIYÖQLIKÖSTNPIÖDINRVELIHFTCS 1251
 Qy 1206 -FRSÖEALÖKÖMHLKKGSGSSTLSRRNKK-----RHFVLÖSKLMTENDSEÖKL 1256
 Db 1252 TEPPTPKLIKYPAAYLK---TTIÖQSDYKSÖVSAQASÖFYLÖRFTL---NGARKÖV 1303
 Qy 1257 KÖTVRRTAKE-----IÖDNT-----KENGID 1279
 Db 1304 PSVTELESITKERNPIFVRITATDGLKGLHIDSATTCÖSSNDLÖSRSPRNVSÖKENGPT 1363
 Qy 1280 IIMADRTFHLIÖSPEDASÖQFVSÖV---HASTDÖEIOH-----MHÖEQANQ 1326
 Db 1364 II---ESFPGIERDLAPTDKLDVLSKVENLÖATLSKÖVNFKPYFKKCLFPD---NTT 1417
 Qy 1327 NAVGTLDVG-----LIDSVCASDSPRNSFVITANRVLHCNADTPEMHMT-- 1375
 Db 1418 NNVPPTISINVENBFYHÖLFNDLFNSNYCÖKÖDÖYÖISIGSLKÖFESSDÖYTDÖIRAMLPG 1477
 Qy 1376 ---TILÖRSKÖDTRVEÖEÖFTVRGMLKEV---KNSPYKSSSLKÖKKR-----WVY 1419
 Db 1478 NGRGKYPTTIDÖKNRFD-----DFTNKYSKHGKSPEÖAKÖMÖVÖLLEKHPLANCÖLIV 1531
 Qy 1420 LTH--NSLÖDYKSSEKÖALKÖTLVNSLCSVVPPEKIFÖKETGYNNVTVYGRKHCYRLY 1477
 Db 1532 CEHÖSSSLPYPKNF-----VLÖLNNÖGNIYDPAÖSKÖLESYKYN-----ÖQÖ 1575
 Qy 1478 TYLÖNÖVTRWSSVIONVTDKAPIDPTÖOLÖIDIKENCLN-----SDV 1521
 Db 1576 QNLKSDKÖSVSIILENKÖSTLÖA-FTGDVÖQLVLSIKESYLRYLNNÖKÖVÖRÖALÖKÖVNSDPT 1634
 Qy 1522 VÖQIYKRNPEILKRY----- 1535
 Db 1635 SLPPFGKNDITITTFKÖQENKMFÖGÖLNGKSGSPVDHVEILSDVPPÖVHPVATISLP 1694
 Qy 1536 ----- 1535
 Db 1695 PMSPTIÖNTNTPPPPSISDSMSPPÖVQÖMLPÖPPPSVSGÖTKRIEITPSLÖIÖPPPS 1754
 Qy 1536 -----HHPLHSPL--LPLPYGDINLNLKÖKGÖYTT----- 1563

Db 1755 SSNSVNSPIGSPMGCIPEPPPTITSVHSLNSGNSPTPPPLPSLSTPTSTPPISSP 1814
 Qy 1564 ----- 1563
 Db 1815 PNFSSSLRVSNLNTNSÖGÖDSSDDPSKRLTVSPALIGTÖSÖLQÖASÖTRPSRÖASTLN 1874
 Qy 1564 -----LÖDEAIKIFNSLÖÖLESMSD-PIP- 1586
 Db 1875 ÖQÖATLKRÖAPVDPTAFYFNKDPKESLIEÖEAKLSKÖKAIKNFSEIÖMM--MGDYPPIPK 1932
 Qy 1587 -----IÖGILÖQHDRLPRLDEIYÖQLIÖKÖTNKVPHPGÖVGNLYSÖQIILTCSCTFLP 1640
 Db 1933 GÖTASLVIÖSTIÖSRIEHNEHRLDEIYÖQÖYÖRÖTNPNPÖVESÖÖKK--GÖELIYPLSITSP 1990
 Qy 1641 SRGILK-YLKFHLKR---IREQPPG-----TEMÖKVALFTYESLÖKTKCEFEVPSR 1687
 Db 1991 SDSLÖQPMÖQÖMSNINIOSSSPÖLÖSLAVICIEKLESHPISÖQ-----RÖMGÖSA 2044
 Qy 1688 DEIEALHÖÖMTSTVYCHGGSCKI-----TINSHTAGÖVEVÖKLIRGLAM-ED 1736
 Db 2045 TEI-----ÖSFRSÖNLÖNGDISTÖKIRFIDÖSTKLAKINTYTTIREINDVÖCRÖYGISÖQ 2098
 Qy 1737 SRNMFALPEYN--GHVÖKAIÖSRÖTVNADVLAKFEKLATSÖVGDLPKPYPKLYCFÖLDT 1794
 Db 2099 SIKÖMGISAVNETAGISÖVÖSETDÖIYDLÖRWEÖ---SEÖKGR---FYÖVÖRRRFPÖLD 2151
 Qy 1795 NVPK-----DSÖEFAFÖEÖAÖEAVIHG--HHPÖPEÖNTÖVÖLÖALRLÖYÖGÖYTL 1843
 Db 2152 DVNKLÖDÖEHLMTDÖDICEFELTYÖQIRÖEMMKGLYTNNEÖDÖSIIÖMILÖLÖLYPNÖS- 2210
 Qy 1844 HÖAIPELÖEYVSLÖRLKARISÖSTKTPPCERLEKRRITSFLEGTLRÖSRFÖGSVÖRÖÖVE 1903
 Db 2211 -----KVLÖFÖKÖVÖRÖYLP 2224
 Qy 1904 BEÖMÖDMITKEÖVSÖSÖRÖSIÖDKRÖKFPÖGÖMÖQÖMÖKYM--ALIKXMPGÖSTLPÖVE 1960
 Db 2225 DÖILNSÖNIÖKWISIEÖQI-----FELVÖQTPÖEYLLÖKMTINLIGSISPEFGTLPÖNIQ 2278
 Qy 1961 CKEGFPÖELMIGVÖADAVSYV--KRGÖRPLEVÖFÖEHILSFGAPLÖNTYKIVÖDER-- 2016
 Db 2279 ÖKEN--PRKÖMLAINKÖGÖVSIFDPHTKESKÖFWTÖQISINÖVAF---TÖDTPÖIMÖGNLÖMK 2333
 Qy 2017 --ELLFÖTSEVVDVÖKMKÖYÖIS 2037
 Db 2334 PIKÖFTTDEHSSIASVÖFYSS 2356
 RESULT 5
 Q9V326
 ID ID Q9V326 PRELIMINARY; PRT; 2167 AA.
 AC Q9V326;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE CK gene product.
 GN CK OR CG7595.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydriidae; Drosophilidae; Drosophila.
 OC NCBI TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley B.M.,

RA Beason K.Y., Benos P.V., Bernat B.P., Bhandari D., Bolshakov S.,
RA Borkova, Borchan M.R., Bouck J., Brokerstein P., Brottier P.,
RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foeler C., Gabriellian A.E., Garg N.S., Gelbart M.M., Glaeser K.,
RA Glöckner A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
RA Jatalin M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,
RA Kammel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Milehina N.V., Mobbart C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shee B.C., Sengen-Klamon I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Striplston M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
RN [2].
RP SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RX MEDLINE=99403001; PubMed=10471707;
RA Ashburner M., Misra S., Roote J., Lewis S.E., Blazer R., Davis T.,
RA Doyle C., Galle R., George R., Harris N., Hartzell G., Harvey D.,
RA Hong L., Houston K., Hoskins R., Johnson G., Martin C., Moshrefi A.,
RA Palazzolo M., Reese M.G., Spradling A., Tsang G., Wan K., Whitehaw K.,
RA Celniker S., Rubin G.M.;
RT "An exploration of the sequence of a 2.9-kb region of the genome of
RL *Drosophila melanogaster*: the *Adh* region.";
RL Genetics 153:179-219(1999).
RN [3].
RP SEQUENCE FROM N.A.
RP STRAIN=BERKELEY;
RC Celniker S.E., Abgaryan A., Arcaina T.T., Baxter E., Blazer R.G.,
RA Buehnoiff C., Champagne M., Chavez C., Chew M., Ciesiolka L., Doyle C.M.,
RA Fafan D.E., Galle R., George R.A., Harris N.L., Hoskins R.A.,
RA Houston K.A., Hummachi S.R., Karra K., Kearney L., Kim E., Lee B.,
RA Lewis S., Li P., Lomoten M.A., Mazda P., Moshrefi A.R., Moshrefi M.,
RA Nixon K., Paclet J.M., Pak S., Pfeiffer B., Poon L., Sequella A.,
RA Sehl H., Sait E., Svirskas R.R., Wan K.H., Weinburg T., Zhang R.,
RL Zietan L.L., Rubin G.M.;
RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1. SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC EMBL: AE003646; AAF53435.1;
DR EMBL: AE003412; AAF44915.1; -;
DR HSSP; P08799; IWND.
DR FLYBase; FBgn0000317; Ck.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MYTH4.
DR InterPro; IPR000567; SBP_dac_1.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 4.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MYTH4; 2.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; P000193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00295; B4; 2.
DR SMART; SM00015; IQ; 3.

Query Match	17.2%	Score 1843.5	DB 5	Length 2167
Best Local Similarity	26.4%	Pred. No. 2.2e-88		
Matches 629	Conservative 381	Mismatches 797	Indels 573	Gaps 88
DR SMART; SMO0242; MYSC 1.				
DR SMART; SMO0139; MYTH4 2.				
DR SMART; SMO0326; SH3 1.				
DR PROSITE; PS50057; BAND_41_3 2.				
DR PROSITE; PS50096; IQ 1.				
DR PROSITE; PS01037; SBP_BACTERIAL_1; UNKNOWN_1.				
DR PROSITE; PS50002; SH3 1.				
DR SH3 domain.				
SW SEQUENCE	2167 AA	250307 MW	3C57E34ADD89A42 CRC64	
Query Match	17.2%	Score 1843.5	DB 5	Length 2167
Best Local Similarity	26.4%	Pred. No. 2.2e-88		
Matches 629	Conservative 381	Mismatches 797	Indels 573	Gaps 88
4 QGTRVWLR-ENGQVPTSTVNS---CAEG-IVTFRTDGVQVTVYKOSTITHOKVTAMHPT 57				
6 RGDYIWIIPASGREEDVAIGARVVSABERRIQVRDDDEGVLAEE-----RIKAMHAS 60				
58 NEEGVDMASTLTHGGSIMYNLFQRYKRNQIWTYIGSILASVNEFYPIAGLYEPATWEO 117				
61 SVQGVEDMISLGDHEAGILRNLLIRYKENIITYTGSILVANPEYQ-ILPIYTGDIQL 119				
118 YSRHLGLPHPIPAIAIECRCLMKRDNOCILKGGSGKSTSLIKLFSLVISQO 177				
120 YKERIGLPHPIPAIIGNAAVAHMKRYRQDQCTIVSGSGAKTSTKILIQYLAIS- 177				
178 SLESLKEKTSCEVAELILESSPIMEAFGNATVYNNSSRFKFFQNLNICQKNIQCGRI 237				
178 -----GKSHWIEQIILEANPILEAFGNATIRBDSNRFEKYIDIHFSANGVIEGATI 230				
238 VDCILSSQNRVVRQNGERNHYIPALLAGLEHERREEFYLSTPENHYALNOSGCEYKT 297				
231 EQYLL-EKSRIVSQNSERNHYVFCYILLAGLSABEKSLDLGMADYKYLTCGNSITCG 289				
298 ISDPSPFEVITAMDVMOFSKEEYREVBRLLAGILHLNIEPITA-----GGAQVSFKTA 352				
290 RDDAEFSDISAKVLLFSDQELWEIKLLAALLHCNIIKKAIVDNLDTETPEHIN 349				
353 LGRSAELLGIDPTQLTDLTQSRMFLRGEELITPLNQAQVDSRDSLAMALYACCEEWI 412				
350 VERVAGLLGLIQLPILDLTRRTLFAHGETVASTSRQSVDRDAFVGIQGRMEVHI 409				
413 KKISRITKGNDFK-----SIGLIDFGFENPEVNHPEQININAYEKLOEYFNKHF 465				
410 KKITVAL-----FKRGTSRNAIGLDFGEPENFQNSFEOCINAYENMLQDFQHL 464				
466 SLEOLEYREGIIVVEDIDMINDGECLELDIE-KGLGLALLIMESHFPQATDSTLLEKHS 524				
465 KLEGEVYHVAIINQHIEFVNDQALDLIAIKQNLIMALLIDBEARFKGTDTMLAKHK 524				
525 QHANNHFYVKRRVAVN-NFGVAKHYAGEVQYDVRGILEKQRDTFRDDLNLNRESRDFLY 583				
525 THGSHKYNLKPKXSINTSFGLNHFAGVVFYDTRGFLDNKRDITFSDDLVLVSQSTNFKLR 584				
584 DLFEHVSRRNQDPTLKGSKRR--PTVVSQPKVDLSLSMATLSSSNPFPRCIKPKNQ 641				
585 QIFA-----QD-IEKAEIRKRPITLSTQFR-KSLDALMTLSSCQPFIRCIKENL 635				
642 KMPQDPAVVLNQLRYSGLMELTVAIRKAGYAVRRPQDFYKRYKVLNMLALPEDVRGK 701				
636 KKPMMFDRGLCCRQLRYSGMETIRIRAGYPIRHGFREFEYRFLP--GVPPAHRTD 693				
702 C---TSLQLDYDANSSEQVLEKTVLEPESLEQKLEKREBEVSHAAWVIRAHVLGFLAR 758				
694 QQAATSRICAVVLGKSDYQLGHTKFLKDAHDLFLEQRDR-----VLR 738				
759 KQYKVLVCAVYIKQNYAPFLRRRFLKKAAALYFQQLRQIARVYRQLAEKREOE 818				
739 K-----IILQSRIGWYRRRFLPLRAALIVQGFQWGYAKRKRYNN----- 782				
819 EKKQBEKKKREBEERERERREAEALRAO-----QBEETKQOELEALQ 865				
783 -----RVGVNRLQALISRVLSHFRIRLGHVIGLQAHARGVLRREVGHHMAVATKIQ 816				

```

Qy 866 KQKAEALTRLEKOK--ENKOVEELIRLEKEIEDLORMKEQOELSTLEASLOKQJERRDQ 924
Db 837 SHVRNRIAMRRYRKLRLEHKQFAEVLQALK--LEE-----QELHRRGK 878
Qy 925 ELRRLLEECRAAOEFLBSLNDEIDECVRNIERSISGSEFSSBLAESACEKRNFNPS 984
Db 879 HAREIAEQHYR---DLHELLEERIEQOLENRRR---VEVNMNINDANAQKQ----- 924
Qy 985 QYPEREEVEG-----FEADDAFQDSPNPSSEHSGSDORTGIRTSDDSSSEDPYMNPT 1038
Db 925 ---BEVVDGKLVEMFPFLPDSSSDAPTP---HGKRETVF---NDLPHANVNODDI 974
Qy 1039 VVPTSEADSTVLLAPSVODSGLSHNSSESESTYC-----MPQN----- 1077
Db 975 IAPHISEDEEDLSEFKQKPAATYFQGVNNOYAKKALKHLLPLHTQGDOLAQAALMI 1034
Qy 1078 ---AGDLPSPDGDYDQDDYEDGATISGSSVT-----FSNYSGOWSPD--- 1119
Db 1035 TILRFTGDMPEPK---YHTMDRMDTTSVMSKVATATGRNPFIRSKFQEAQLMGLDPAEL 1091
Qy 1120 ---YRCSVGTNSSGAVRFSEGAQSFDESEDPDS-----RPTDDE 1160
Db 1092 KQKPSIRKHLVSLTKRKQKLG---EDVRRRLQDDEYTDASYQSMQSRPISNLEK 1145
Qy 1161 LSY-----RSDSVYS--CVTLPYFHSFLYMKGLMNSWKRM-----CVLKD 1200
Db 1146 LHPITIGHGLRAELREIYCOJCKOL-----TNPLK3SIRAKRMILLSCVCGFAPS 1197
Qy 1201 EYFL--MFRS-----KQBALKQGLMHKKGGSGSTLSRNNKKRMFVLRQSK--- 1244
Db 1198 EKFVNYLRAFIREGPPGVAFYCEERLKRTPFN---GTRNOPSPMLEIQATKSKK 1248
Qy 1245 ---LMYENNSEBKLKGTVEVRTAKEIINDNT-----KENGIDIIIMADRTFHIIAS 1293
Db 1249 PMLPTTFMDGNTKTLADS--ATTARELCNOLSDKISLKQGFSLYA--LFDKVSLL 1304
Qy 1294 PBDASQMFSLSQVNA--STDQEIQ-----MHDEQAMP--QNAV 1329
Db 1305 GSGGDHVMALISQCEBYAKEQGOENNAWRLEFRKEIFAPMHEPRHDQVNTLIQOYV 1364
Qy 1330 GTLDVGLIDVSCASDPDRPNSFVIITANRVLHCNADPEIMHMHTLLQRSKQDTRVGG 1389
Db 1365 RGCVKFG--EYRC-----DKSEBDLAMIAOQYF--IBYSTDMEERLFTLLPNFIPDCLSG 1416
Qy 1390 QSEIYRGW-----LHKEVKSPPKSSSLKTKK-----RMFVTHNSLDYKSEKN 1434
Db 1417 VKRALERMALVLOAYKKSYYVKDKIAPLKIKEDIYSVAKYKMPLLFSRYEAYRNSGPN 1476
Qy 1435 ALKLGTLVNLNLSCS--VPPDEKIFKETGYMNTVYGRKHCYVLYTKLNEAT----- 1485
Db 1477 LRKNVDYIAVNTGVVVDQBOVLELSEFPEITANVSQKTNKVFTQTSLSLVRESEPT 1536
Qy 1486 ---RMSVSIQVNTDTKADIDTPJ-----QOLIDIKENCIN 1518
Db 1537 FQSPNADIRDLVYVFLDGLKRSKSKYVIALQDYRAPSDDGTSFLSPFKGLIILEDSCGE 1596
Qy 1519 SDV-----VEQIKRNPILRYTHHPLSLPLPYGDJ-----NINLAK 1557
Db 1597 SVLNNGMCTIGRCDSOGRDPPAETVYV--PPLSKRPDIDIALFNIIEAHNGRRLSMS 1654
Qy 1558 DKGYT-----TLQDEAIKIFNSLOQLSMS-----DPI--PIIO- 1589
Db 1655 NGGAVPRDRPHLMYALDHR--LPPKRTMSKITLSSKRSSELMRYSRDPIKAPLLK 1713
Qy 1590 ---GIIQTGHDL--RP-----LRDELYQOLIK 1611
Db 1714 LOSKEFAEACPAFAALIKYMGDLDSKPRMGNEITDHI FDGPLKGBILIRDEIYQOLMK 1773
Qy 1612 QTNKVHPSPGVGLYMOILITGSCFLPSRGILTKLXKHLKXIRQFQGTBEKALFT 1671
Db 1774 QL--TDNRMRSEERGEMLMWATGLFACSGGLKELLLPLRTRRRPISODSMHR----- 1826

```

```

Qy 1672 YESLKT--KREFVPSRDEIALIHR--QEWTSITYCHGGGSKITINSHTTAGEVVEK 1727
Db 1827 ---LQKTIIRHGQRKYPHQVEVEALQHTQGLFHKVYPPDDTDEAFVDSSTRADFCNN 1883
Qy 1728 LIRGLAMDSRMPL-----FEYNHVDKAYSRIYVADVLAKEFL 1770
Db 1884 ISQRLSTLSTSEG--FSLFKVIADKVISVECGFFDFVHLLTDWIKKARPDR----- 1934
Qy 1771 AATSEVGLPMKFYKLYCF--LQTDNVP--KD--SVEFAFMFEQAHEAVIHGHHPAPENL 1826
Db 1935 ---GANP--QFTYQVFPMKLMTNVTYPGADRNDLFIHQBELPKILRGTHKSRBEA 1987
Qy 1827 QVLAALRLQYLOGDYTLHAIPLEEVYSLORLKARISQSTKTPTPCERLKRRTSPLEG 1886
Db 1988 AKTALVFRVRGE-----NKQELQ--IQMLRELIPSDIMKIQSTS----- 2028
Qy 1887 TLRBSFRIGSVRVQVREBQMLDMYIKESVARSASIIDKRRKFGGQAOQOMAKYMALI 1946
Db 2029 ---EW-----KRSIVASYNQDGGMTSEDAKVAFLKIV 2057
Qy 1947 KEMPGVSTLPDV--ECKEGFPQBLMGVADAVVYKREGRPLEVQYEHILSFGAPL 2005
Db 2058 YNMPFTGSAFPEVKQTTBPNTPEMLLIINKHGVSLIHPVTQDILVTBPFTRISSWS-- 2115
Qy 2006 ANTY-----KIYDERELLFETSEVVVDVAKLMKAYISMIV 2040
Db 2116 GNTVFPMTIGNLVRGSKLLCETSLGYKMDLLTSTISLML 2155

```

RESULT 6

```

ID Q9DGG8 PRELIMINARY; PRT; 2179 AA.
AC Q9DGG8;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DE 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Myosin VIIA.
GN MYO7A.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_TaxId=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TUBERINGEN;
RX MEDLINE=20414642; PubMed=10958658;
RA Ernest S., Rauch G.U., Haeflter P., Geisler R., Petlic C., Nicolson T.,
RT "Mariner is defective in myosin VIIA: a zebrafish model for human
hereditary deafness.";
RL Hum. Mol. Genet. 9:2189-2196 (2000).
DR EMBL, AJ404002; CAC05419.1; -.
DR HSSP, P08799; 1MND.
DR InterPro, IPR000299; Band_4.1.
DR InterPro, IPR000048; IQ_region.
DR InterPro, IPR001609; myosin_head.
DR InterPro, IPR000857; MYTH4.
DR InterPro, IPR001452; SH3.
DR Pfam, PF00373; Band_41; 1.
DR Pfam, PF00612; IQ_5.
DR Pfam, PF00063; myosin_head; 1.
DR Pfam, PF00784; MYTH4; 2.
DR Pfam, PF00018; SH3; 1.
DR PRINTS, PR00193; MYOSINHEAVY.
DR ProDom, PD000355; myosin_head; 1.
DR SMART, SM00295; B41; 2.
DR SMART, SM00015; IQ; 4.
DR SMART, SM00242; MYSC; 1.
DR SMART, SM00139; MYTH4; 2.
DR SMART, SM00326; SH3; 1.
DR PROSITE, PS00057; BAND_41_3; 2.
DR PROSITE, PS00096; IQ; 2.
DR PROSITE, PS00002; SH3; 1.

```

SEQUENCE 2179 AA; 251655 MM; FPA40D1EFD588P76 CRC64;
Query Match 17.1%; Score 1834.5; DB 13; Length 2179;
Best Local Similarity 25.7%; Pred. No. 6.5e-88;
Matches 624; Conservative 389; Mismatches 780; Indels 639; Gaps 83;
4 OCTRWVL-RENGQHPSTVNSCAEIVFRTDYGOVFTYKQSTITTH-----QKTYAMH 55
6 OGVOYVLDLKTGHFEFVPI-----GAVVKLCDSGQQLVLDDEGNEHMSIPQNTNITKPMH 60
56 PTNEBEGVDMASTELHGGSIMTNLPOKYKNOIWTYIGSLASVNPYQPIAGLYEPATM 115
61 PTSHGEEDMIRGDLNEAGILRNLLIRREHLIYTYGSLVAVNPYQ-LLEPIYADQI 119
116 EGYSRHLELPHIFALANECYRCCLKRHNDQCILIKGSGAGKTESTLILKPLSVIS 175
120 RLTYNKKIGEMPHIFALADNCTFNNORNNKQDCIISGSGAGKTESTLILQPLAAS 179
176 QOGLLELLEKETSCEVERALLESSPIMEAFGNAKTVNNNSRFGKPVOLNICOKGNIQGG 235
180 GQH-----SWIEQVLEANPILAEAFGNAKTINDNSRFGKVIDHFNKRGALIEGA 230
236 RIYDCLLSQNRVVRQNGERNHYITALLAGLEHEBEREFYLTSPENTHYLNQSCVED 295
231 KLEQYLL-EKSRVCRAQDERNHYIFCYMLKGMTPOKQGLSKATDYTLITGCTYC 289
296 KTISDOESPREVITAMDVQFSKEVEVRSLLAGILHGNIE-----FITAGAGVQSFK 350
290 DGDHDDQKEYSNINSAKVLMTFDKENWEISKLLALILHMGNLRYEARTYNLDACEVRC 349
351 TALGRSAELLIDPTQITDALTORSMFLRGEELITPLNVOQAVDSRDSLAMALYACCFEM 410
350 SALTAAVLELVLDKDLNCLTSRTIITRGETVSTPLISIQALDVDAFVKGIGYRLFVW 409
411 VIKKISRIKGNDF-----KSTIGLIDTFGEFENFVNHEQONINAYNEKLOEYENKHI 464
410 IYKINAAIYKPSLELKAVERSIIGLIDTFGEFENFVNHEQONINAYNEKLOEYENKHI 469
465 FSLQLEYSREGVWEDIDWIDNGECLDLE-KKLGLALINEESHPOATDSTLEKHI 523
470 FKLQSEYENLENNMOHIEFTDNOALDMALIKPMNIIISLIDESKFPKKTDTTMLANKLN 529
524 SOHANNHFYVKRP-VAVNNGVKGIVAGEVOYDVYRGILEKNRDTFRDILLNLRESRFDI 582
530 SQKHLNTNYIPKQVETETGFIQHPAGVYVYETRGLEKRDITLHGDIIDLVHNSKXKFI 589
583 YDLFEVHSSNNODTLKCGSKHRR--PTYSQFKVDSLSHMAITSSNPFRCIKPMN 640
590 KQIF-----QADVAMGAEIKRSPTLSSQFK-RSLLEMLRTLSVCOFPFRICIKPMN 640
641 OKRPDQFQAVLNQLRYSGLMETVRIRKAGYAVRRPQDFYKRYKVLAMNLA---LPED 697
641 YKRPMLFDRCLCYRQLRYSGLMETVIRKAGYPIRTTFVAVRYKVLAMNLA---LPED 700
698 VRGKCTSLQLYDASNSEWOLGKTVFLRESLEQKLEKRESEVSHAANVIRAVLGFLA 757
701 LRGTGCRIRAEAVLGRDDDMQMGKTKIFLKHDMMLLEIRDKAITDKVILIQVAVGFKD 760
758 RKQYKRVLYCVVLIQ-----KXY----- 775
761 RSNFLMKKKSAMLIQKTRGVCYKNGYAGNRGFSRLQALYRSRKLQYTVHVARQIMLF 820
776 ----RAFLRRRFLHAKKAIIVEFOKQROGQIARVYRQLAAE--KREOEKKKQOESEK- 828
821 QGRGRLVRRARHRLMAVITTOATYTRGMIAARLTKRLKGEYRRRLIEAKKFLABEOUKI 880
829 ----KKREBERERERERAEAL-----RAOESEETRKQOELEALOKSOKEA---- 871
881 RNQMSARKAKEBAEKHGERLAQIAREDAERKEKQGEARMKEMLDQMEKARQEPVNS 940
872 ----ELTRELEKQKQKQVEIILRL-EKEIEDIQRMK 903
941 DNVDMKFGFLGTTNSPFGQGOAPAGFEDLERTHRELEVDDLESLPLPEDDLEDSBYK 1000

904 -----EQEELS----- 909
1001 FAKFATYRQGTSTHYMRRLKQPLRFHEDGDDQALAAVAVITLRFMGDLPEPKYHTA 1060
910 -----LREASIQLOERRDQELR-----RLEBACRAAOEFLSLNFDID 950
1061 ISDGSEKIPVMTKIYETLTKTKYKRELQALQEGEGENTNIESHKSSVRRKVLSTLKKKS 1120
951 ECVRNIERLSGSGSEFSSSLASACEKKNFNS-----QPYEESEVDGFEAD 999
1121 KTBEEVTKRLNG-EYVH-GNSMLEDRPTSLKXHLFTGNGILRPGLRDI---YCOI 1175
1000 DDAFKDSPNSRSG-----HSDOR----- 1018
1176 CQOLQNPBSKSHAGWILMSLCVCGFAPSEKFKYKLRNFISEGPGVAPYCEERLRRF 1235
1019 TSGIRTSDDS-----SEEDPYM-----NDTVPTSPSADSVLLAPSVQDSGLHN 1064
1236 VNGTRTQPSMLELQATSKKPIMLPVTFMDGTTTLTDSATYAKELCNALSDKISLQD 1295
1065 -----SSGSESTYCMQNGDLRSPDGDYVDQDDYEDGAITSGSSVTFEN 1110
1296 RGFSLYALFDKVSLSGSGN---DHVMDAVSQCEQYAKKEGAORNA---PWRLFPRK 1348
1111 SYGQSPDPYRCVGT---YN-----SSGAYRFSSEG-AQSSFESEDFSRPDTDE 1160
1349 ELFTPMHDAEBAVATNLVYQIVRGVKGGEYRCRQEDLAELASQYVYDYGSELIVERL 1408
1161 L-----SYRRD-SVYGCVTLPYHSFLY-----MKGLANSWKRRWCVL 1198
1409 LELISPYIDRDISAKTYERWAQFIAMAKKGVYTOKKVDQKYEEVLEDFARXKMPLL 1468
1199 KQETLMPRSKQBALKQGLHKKGGSSSTLSRNKKRMFVLROS---KLWYFENDSE 1254
1469 FRRFYEAFFSGSPSLPK-----NDYIVAVNNTGVFVDEQOVLLLELSPETLAAS 1519
1255 KLGK-----TVEVTAKEIINDTTKENGIDIDIMDRFPHLAESEPDASQWFVLSQVH 1308
1520 SSKGGLQAGSTLNTATIKADEFTFSNAMEDRLVYTF--LEGLRKRSKRVALL-- 1572
1309 ASTDQEIQEMHQANPONAVGTLDVGLIDVSCASDSPRPNFVIITANRYLHCNADTP 1368
1573 -----QONP-----SPAADS----- 1583
1369 EEMHHITLQKSKDTRY---EGOEPIVRGWLKXVNSPKMSLKKRW----- 1417
1584 -----TFLSLKGDVLVLDQDTEBQVNTSGMAHG-----TNDRTKQGRDPADCV 1628
1629 YVLP-----TVRRPHDV---VAVVTMPRQROSLAR- 1658
1478 TKLNEATRMSVIONVDTKAPIDTPQO-----LIQIKCNCLNSVDVEQIYKRNPL 1532
1659 -----SHVALTEETEEVRVRYLLEFSYDYFRPPRPGLTSRWMITQNRGDKLM 1707
1533 RYTHHRLSPILRLPYGQINLNLKDKGYTLTQDAIKIFNSL-----QOLES 1580
1708 CCTREPIKALK-----KCGHEELSQEACAFIVMKYMGDYPSKRTSRVNE 1756
1581 MSDPIIIOGILQTHDLRLPRLDELYCOLIKO--TNKYRPHSPSVGNLSMOJLTCLSCFP 1638
1757 LTD--QIFEGALKA-----EPLKDELYCOLIKQJLTENHIIKYSEKQ---WELLMVCVGF 1806
1639 LPKSLGLYKFLKRLRQRFQTEMEKXALFTYESLKK--TKREVPYSDIEBALIH 1695
1807 PPSNVLPHVQRFLOS-KGHP-----LALQCMQRLQALRNGSRKYRPHLVEBALOH 1859
1696 R-QEMTSTYVCHGGSCKTITNSHTAGB---VVEKILRGJLAMDSDRMFA----- 1742
1860 KTTQIFHKVYPPDDTDEAEVSSSTKADFCINTISSRL--LKTDEGFLFVKISDKVIS 1917


```

OY 1743 -----LPEYNGHDKALSESTVADVADLAKREKLAATSEVODLPKKEFF--KLYCFDLD 1794
Db 1918 VPBEDFFDFPRHLIDWIKKARPAKXGI-----VSLTYQVEFMKLM-----TT 1962
OY 1795 NVP-KDS-VEFAFMEQAHAVIHGHHPAPEENIOVLAAL--PLQYLOGDYTLMAAIPPL 1850
Db 1963 TVPQKUSPADBIFFHYOELPKTLRGYKHCSEBEVFOLALYYKXF-EDDSHSPSIPKM 2021
OY 1851 BEVYSIORIKARISOSTKTFPTPCERLEKRRTSFLEGLTRBSFRGTSVVRQKVEBOMLD 1910
Db 2022 -----LKEMIPQ-----DLRQLSPDD----- 2038
OY 1911 WIKEEVSSARASIIDKMRKFOQMNOEQAMAYMALIKEMFGXSTLFDV-ECKEGEPQOE 1969
Db 2039 W-----KRSIYAVFMRHAGKCSREBAKLMFLKIIFFMFFGSAFFEVQKPTTBHPZEI 2090
OY 1970 LMLGVSAIDAVSVYKRGEGRPLEVFQYEHILSPGAPLANTYKIV-----VDERELLFETSE 2024
Db 2091 LLIIINKGVSLIDPKNKDILITTPYPTKISWMS--GNTYFHITIGNLVQSCKLCESTLS 2148
OY 2025 VVDVAKMKAYIISMLVKKRYSTTSASQSGSS 2056
Db 2149 GYKMDDLTSTYSIQMLT--TWSKORNSRGSS 2177

```

Query Match	Score	DB ID	Length
17.1%	1831.5	DB 1	2179
25.7%	Pred. No. 9.3e-88		

Matches	624;	Conservative	386;	Mismatches	784;	Indels	637;	Gaps	82;																																																																																																																																																																																																																																																																																																																																																																																										
QY	4	QGTIVWLI-RENGQHP	PSTVNSCA	GIVFRTDYG	VFYTKOSTIT	-----	OKVTAMH	55																																																																																																																																																																																																																																																																																																																																																																																											
Db	6	QGDYVWLDLTKGHEFVPI	-----	GAVVKLCD	SGOIQVLD	DEGNEH	ISPQNA	TIKEMH 60																																																																																																																																																																																																																																																																																																																																																																																											
QY	56	PTNEEGVDMA	SLTE	HGGSIMYL	LFQRYK	NO	WTYIG	SLASVNYOP	AGLYERPTM 115																																																																																																																																																																																																																																																																																																																																																																																										
Db	61	PTSHGVEDM	RLADL	NEAGILRL	LR	YREBHL	LYTYTGS	LVANVYO-LLPYTDQI	119																																																																																																																																																																																																																																																																																																																																																																																										
QY	116	EQYSRRLGELPHI	PAI	NECVY	CLMK	RHNO	CIILK	SGESAG	TESTKULIFELVIS 175																																																																																																																																																																																																																																																																																																																																																																																										
Db	120	RLTYNKCI	GEFPHI	PAI	ADCTY	YNNQ	RNNK	DOCCIS	SGESGAK	TESTKULIDPLLAIS 179																																																																																																																																																																																																																																																																																																																																																																																									
QY	176	QOSLELS	KEKTS	CEVERA	LES	SPIME	AFNA	KTVNN	NSRFPKPOL	INICQKNIQGG 235																																																																																																																																																																																																																																																																																																																																																																																									
Db	180	QGH-----	SWIEQV	LEAN	PILE	AFNA	KTI	RNDNS	RFPKTYID	HNKKGALIEGA 230																																																																																																																																																																																																																																																																																																																																																																																									
QY	236	RIYDCI	ISSQ	NRVVR	QNG	PERN	YHIF	PALLA	GLE	HEHER	EEFYLS	TENYH	YNOS	GEVD 295																																																																																																																																																																																																																																																																																																																																																																																					
Db	231	KIEQYLL-EX	RYCQ	RARD	ERNYH	I	FCY	MLK	GMT	PDQK	QOLG	LSK	ADITYLL	TIG	NTVC 289																																																																																																																																																																																																																																																																																																																																																																																				
QY	296	KTID	QSF	FRVIT	IA	MO	QSE	KV	EV	SV	SLA	GI	HL	GNIE	-----	FTAGA	QVSPK 350																																																																																																																																																																																																																																																																																																																																																																																		
Db	290	DGRDDQ	KYS	NI	RAM	KV	ML	FT	DE	NE	IS	TL	LA	IL	MG	LR	YEATYD	NLDACE	VVRC 349																																																																																																																																																																																																																																																																																																																																																																																
QY	351	TALGR	SAB	LGLD	P	TO	LT	AL	TQ	SR	MF	L	R	GE	IL	TP	LN	VO	Q	AV	S	R	S	L	A	M	A	L	A	C	F	E	W 410																																																																																																																																																																																																																																																																																																																																																																		
Db	350	SALT	TA	V	L	E	V	D	L	K	D	L	M	N	C	L	T	S	T	I	T	T	R	G	S	T	V	E	P	L	S	T	E	O	A	L	D	V	R	A	F	V	K	I	G	R	L	E	W 409																																																																																																																																																																																																																																																																																																																																																		
QY	411	VIKKINS	R	I	K	N	E	D	E	-----	KS	I	G	I	L	I	F	G	F	E	N	H	E	Q	F	N	I	N	A	N	E	K	L	O	E	Y	F	N	K	H	I 464																																																																																																																																																																																																																																																																																																																																																										
Db	410	I	V	E	K	I	N	A	I	A	Y	K	P	S	E	L	E	K	A	V	A	R	S	I	G	L	D	I	F	G	E	F	E	N	H	S	C	I	N	E	N	E	L	O	F	V	A	R	H	V 469																																																																																																																																																																																																																																																																																																																																																	
QY	465	FSL	E	Q	L	S	Y	S	R	G	L	W	E	D	I	M	I	D	N	E	C	D	L	E	-----	K	G	L	L	A	L	I	N	E	S	H	P	O	A	T	S	L	L	E	K	H 523																																																																																																																																																																																																																																																																																																																																																					
Db	470	F	K	L	E	O	E	E	N	E	N	I	M	O	H	I	E	F	T	D	O	A	L	M	I	A	I	K	M	N	I	S	L	D	E	S	K	P	K	G	I	D	T	M	L	K	N 529																																																																																																																																																																																																																																																																																																																																																				
QY	524	SO	H	A	N	N	H	F	Y	V	E	R	-----	V	A	N	N	F	G	V	K	I	A	G	E	V	O	D	V	G	I	L	E	K	R	D	T	R	D	L	L	N	L	R	S	R	P	D	F	I 582																																																																																																																																																																																																																																																																																																																																																	
Db	530	S	O	H	L	T	N	N	I	P	P	K	T	Y	E	T	O	F	O	I	G	F	A	G	V	Y	E	T	R	G	L	E	K	R	D	T	L	H	G	I	I	O	L	V	H	S	K	K	F	I 589																																																																																																																																																																																																																																																																																																																																																	
QY	583	Y	D	L	F	E	H	V	S	S	R	N	O	D	T	L	C	G	S	K	H	R	-----	P	T	V	S	O	F	K	V	D	S	L	H	S	M	A	T	L	S	S	N	P	F	V	A	C	I	K	N	M 640																																																																																																																																																																																																																																																																																																																																															
Db	590	K	O	I	F	-----	Q	A	D	V	A	M	G	A	T	R	K	S	P	L	S	S	O	F	-----	R	S	L	E	L	M	R	T	S	I	C	O	P	F	V	A	C	I	K	N	E 640																																																																																																																																																																																																																																																																																																																																																					
QY	641	Q	K	M	D	O	P	O	A	V	L	N	O	L	A	I	S	G	M	L	E	V	-----	I	R	K	A	G	I	A	V	R	P	O	F	Y	K	I	K	L	M	R	L	A	-----	L	P	E	D 697																																																																																																																																																																																																																																																																																																																																																		
Db	641	Y	K	K	M	L	P	D	R	E	L	C	V	O	L	T	Y	S	G	M	M	E	T	I	R	R	A	G	Y	P	I	R	T	F	V	E	V	D	V	R	Y	L	M	P	V	K	E	A	Y	K	O	E	D 700																																																																																																																																																																																																																																																																																																																																														
QY	698	V	R	G	E	S	T	L	L	O	Y	D	A	S	N	S	E	M	O	L	G	K	T	V	A	L	R	S	L	E	O	K	E	R	E	E	B	V	S	H	A	M	V	R	A	N	V	L	E	P	L	A 757																																																																																																																																																																																																																																																																																																																																															
Db	701	L	R	G	T	O	R	I	A	E	V	L	G	R	D	D	Q	M	O	G	K	T	I	P	L	K	D	H	M	D	L	E	I	E	R	D	K	A	I	T	D	V	I	L	I	O	K	V	A	R	G	P	K	D 760																																																																																																																																																																																																																																																																																																																																													
QY	758	R	K	O	T	R	K	L	Y	C	V	A	I	O	-----	K	N	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	

```

Qy 910 -----LTFASLOKLOERRDOELR-----RLEEBACRAOEFLESINFEID 950
Db 1061 ISDSEKIPVMTIYETLGKTKYRELQALQGEENTHIESHKSSVRHVLVLTLLKKS 1120
Qy 951 ECVRNIERLSIGSGSFSSSBLAESACEKNFNFS-----OPYEEVEDEGEAD 999
Db 1121 KITEEYTKRLNDG-EYTVH-GNSMLDRPTSNEKLHFIINGILRGLADEI---YCOI 1175
Qy 1000 DDAFQSPNPSSEB-----HSPOR----- 1018
Db 1176 CQOLNPNPSKSHARGWILMSLCVCGPASEKVFYKLRNFISGPPGVAFCERLRRTF 1235
Qy 1019 TSGIRTSDDS-----SEEDPYM-----NDTVPTSPSADSTVLLAPVODSGSLN 1064
Db 1236 VNGTRQOPSPWELQATKSKKPIMLPVTMDDGTTKTLTDSATTAELCALSDKSLQD 1295
Qy 1065 -----SSGSESTYCMFONAGDLPSPDGDYDQDDYDGAITSSSVTFSN 1110
Db 1296 RFGFSLYIALFDKYSSLSGSGN-----DHVMDAVSQCEQYAKEQQAQRNNA---PWRLFFRK 1348
Qy 1111 SYGSGSPDYRCVGT---YN-----SSGAYRFSSEB-AQSSFDESEEDFDRPPTDDE 1160
Db 1349 EITFPHNDPAEDVATNLITQOIVRGVKEGYCDREDLARLASQOYVYDGSSEILVERL 1408
Qy 1161 L-----SYRSD-SVYSCVTLFPYHSPLY-----NRGGLMNSKRRMVCVL 1198
Db 1409 LSIIPSYIDPRELSSAKTYERMAQFIAMAHKKGVYQKVDPOKVEEVVDFAFRYMPLL 1468
Qy 1199 KDETFLMPSKQBALKQGMHLHKGSGSTLSRRNKKRMFVLQOS---KLMYFENDSEB 1254
Db 1469 FSRFYEAFFKSPSLPK-----NDVVAANWMTGVFVDEQEOVLLELSPETAVS 1519
Qy 1255 KIKG-----TVEVFAKEIINDTKENGIDIMADRTFLIASEBDSQMSVSLSQVH 1308
Db 1520 SSKGKQLAQOSFTLATITKADEFPTSNMEDIRDLVTF---DEGLRKSKEFVAL--- 1572
Qy 1309 ASTDOEIOEMHDOANPONAVGTLVGLIDSVCASDSPRPNFVITIANRVLHCHADRP 1368
Db 1573 -----QDNP-----SPAADS----- 1583
Qy 1369 EEMHMITTLQNSKQDTRV---EGGEFIVRGWHLKEVKNSPMSSKLKGRW----- 1417
Db 1584 -----TFLSPKGLDVLVDQDTGEQWMTSGWAHG-----TMDRTQORDDPPADCV 1628
Qy 1418 FVLTNHSLOYKSSSEKNAKLGTLVNSLCSVVRPEBKIFKENGYNVNVYVGGKHCYRL 1477
Db 1629 YVLP-----TVVRPHDV---VAFVTMTPDQOESLRT- 1658
Qy 1478 TKLNEATWSSVIVQNTDTKARIDTPTOO-----LIQIKENCLNSDVVEQIYKKNPIL 1532
Db 1659 -----SHVPALTEEBERVKRYTLEEFSDYDFRPPKPTLSRWITAKRGDKLM 1707
Qy 1533 RYTHNPLHSPLELPYGDINLNLKDKGYTTLQDEAIKIFNSI-----QOLBS 1580
Db 1708 CCTREPIKQALLK-----KCGHEHLSQEAQMAFIAVMKMGDVPKRTSRVNE 1756
Qy 1581 MSPRILIGLITGHDRLRLDELYCOLIKO---TKVHPRGSVGNLYSMQITLCSCTF 1638
Db 1757 LTD-OIFEGALVA---BPKOEIYCOLIKOLTEHNIKSEBK---WELLMLCVGLF 1806
Qy 1639 LPSRGLIKLKFHLKRIEQFPCTEMEKVALFTYESLKK---TKCREPAPSRDEIALH 1695
Db 1807 PRSNVLLPHVQRLQS-KKHNP-----LALDMQQLQALARKSGSKYRPHVLVEVRIQH 1859
Qy 1696 R-OEMTSTVYCHGSGCKITINSHTTAGEVEVELIRGLAMEDSRNMFALF----- 1744
Db 1860 KTIQIFHKVYFPDDTBEAFESSTAKFCCLNISRLILKTPEG-FSLFVKISDKIVSI 1918
Qy 1745 -----EVNGHDXAIESKTVADVLAKEKLAISVQDLPWKRYF---KLYCFDJDON 1795
Db 1919 PBGDPLDFVRHLTDMIKKARPAKDG-----VPSLTYQVFMKMLM---TTT 1963

```

```

Qy 1796 VP-KOS-VEFAFMFOAHNAVIGHHPAPEENLOVLAAL--RLQYLOQDYTLHAALPILE 1851
Db 1964 VEGKDSFADISFIHYQELPKYLRGYHKCSREVEVQFGALITYVKE-EDDKSHFPIPKM- 2021
Qy 1852 EYVSLQRLKARISQSTKTTTPCERLEKRTSLTBGLTRSPRTGSVNVQKVEEQMLDMW 1911
Db 2022 -----LKEMTPO-----DLIRQLSPD-----W 2039
Qy 1912 IKEEVSASASIIDKRRKFOGNORQAMAKYMALIKEMPGYSTLFDV-ECKEGFPQBL 1970
Db 2040 -----KRSIVAFNRNAGSKSREBAKLMFLKIFKMTFSGAFPEVQTEPRHPELL 2091
Qy 1971 WLGVSADAVSYKKEGRLPEVQYEHILSFAPLANTYKIV-----VDERELFETSEV 2025
Db 2092 LIAINKVGSVLDIPKNKDLITTPPTKISMSS--GNTYFHTIGNLVQSGSKLCTSLG 2149
Qy 2026 DVVAKLMKAYISMVYKKRSTRSSSQSS 2056
Db 2150 YKMDLLTSTYSIOMLT--TMSKORNSRGS 2177

RESULT 8
P91443
ID P91443 PRELIMINARY; PRT; 2098 AA.
AC P91443;
DT 01-MAY-1997 (Tremblrel. 03, Created)
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE T10H10.1 protein.
GN T10H10.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berke M.,
RA Bonfield J., Burton J., Connell M., Copeay T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonnenhammer E., Staden R., Suisson J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat U., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans."
RL Nature 368:32-38 (1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Nelson J., Langston Y.,
RT "The sequence of C. elegans cosmid T10H10."
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.,
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL; U08848; AAB37988.1; -.
DR HSSP; P08799; 1MDN.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR000857; MYTH4.
DR InterPro; IPR000159; RA_domain.
DR Pfam; PF00612; IQ_4.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00784; MYTH4; 2.
DR PRINTS; PR00193; MYOSINHEAVY.

```


Qy	1669	EIEALIIH- QEMTS IVYUHGCGGSCKITINSHTTAGEVVEKILRGJLAMSDSNMFLFEYN	1747
Db	1767	EVEALQHKTTQIFHNKFFPPNDTDEALIVDSATRADPCBKIGYRLGLKS--DGSLF---	1822
Qy	1748	GHVDAIESRTVAVDLAKFEKLAATSEVGDLPWKFEYKLYCFELTDNVPKDSVEFAEMF	1807
Db	1823	-----VKIKDKVLA-----VPES--EFFPDY	1841
Qy	1808	EQAHBAVYHGHHPABE-----NIOVLALRLQYLOGDYTLULIIPPLEEYVLSQRLKA	1861
Db	1842	VRSLSDWHTNHAQOKDATTMIPINQVYEMRLTW---NPAQADPADIIFNHYQ---	1894
Qy	1862	RISOSTKFTPCSERLEKERTSFLEGTLSRFFSTGSVABQKVEBEOMLPMWIKBE-----	1915
Db	1885	---ESQKLLGYNHKTQNDVIELAIIIRSTKQSKNAPLQIPQJLBDEIIPKDSLYKS	1951
Qy	1916	VSSARASIIIDKMKRFQGNNOBOMAKYMALIKEMPYGSLTFDV-ECKEGGFPOELMGV	1974
Db	1952	ASEMKRTISNAYARIENHKSPDAKIEFLNYICRMWTFCSAFPPVQYSDMLNPRLLAI	2011
Qy	1975	SADAVSYKKGGRPLEVFOYEHLISFGAPLANTY-KIVV-----DERELFTSEV	2026
Db	2012	NOTGNVNIHDLTKMLLVQPPFVLCNMYS--GNTVFNMVGNMLKNGSKLLDLDTVGX	2063
Qy	2027	DVAKLMKAYISMIYKKRY---STRSAS	2051
Db	2070	KMDLLTYSILLISNONNBPCKIREVA	2097

RA Palenzola M., Piltman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenhach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye Y., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,"
 "The genome sequence of *Drosophila melanogaster*,"
 RT Science 287:2185-2195(2000).
 RL EMBL; AE003618; AAF52536.1; -.
 DR HSSP; P08799; IMND.
 DR FLYBase; FBgn0040299; My2881.
 DR InterPro; IPR0000299; Band_4.1.
 DR InterPro; IPR0000048; IQ_region.
 DR InterPro; IPR001609; myosin_head.
 DR InterPro; IPR000857; MYTH4.
 DR Pfam; PF00612; IQ_3.
 DR Pfam; PF00063; myosin_head; 1.
 DR Pfam; PF00784; MYTH4; 2.
 DR PRINTS; PR00193; MYOSINHEAVY.
 DR ProDom; PD000355; myosin_head; 1.
 DR SMART; SM00295; B4; 2.
 DR SMART; SM00015; IQ_3.
 DR SMART; SM00242; MISC; 1.

RESULT 9			
Q9VLT3	PRELIMINARY;	PRT: 2129	AA.
ID	Q9VLT3.		
AC	Q9VLT3.		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	CG6976 protein.		
GN	MYO28B1 OR CG6976.		
OS	Eukaryota: Melanogaster (Fruit fly).		
OC	Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta;		
OC	Pterygota: Neoptera: Endopterygota; Diptera; Brachycera; Muscomorpha		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_taxid=7227;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	STRAIN=BERKELEY;		
RC	MEDLINE=20196006; PubMed=10731133;		
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,		
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,		
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,		
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,		
RA	Brindon R.C., Rogers J., H.C., Blazek R.G., Champe M., Peiffer B.D.,		
RA	Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,		
RA	Abriil J.F., Agbayanti A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,		
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,		
RA	Beeson K.Y., Benos P.V., Berman J.P., Bhandari P., Bolshakov S.,		
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,		
RA	Burtis K.C., Buesam D.A., Butler H., Cadieux E., Center A., Chandra I.,		
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,		
RA	de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,		
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,		
RA	Doulier K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.		
RA	Fowler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,		
RA	Glodex A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,		
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Howack J.,		
RA	Hoslin D., Houston K.A., Howard T.J., Wei M.-H., Ilegwam C.,		
RA	Jajalji M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,		
RA	Kimmel B.S., Kodira C.D., Kratt C., Kravitz S., Kulp D., Lai Z.,		
RA	Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,		
RA	Li X., Mettel B., McIntosh T.C., Meled M.P., McPherson D.,		
RA	Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,		
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,		
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacib J.M.,		

DR	PROSITE; PS50096; IO; 2.	2129 AA; 24414 MW; 1A90EBDC35103481 CRC64;
SR	SEQUENCE	
QY	Query Match	15.9%; Score 1705; DB 5; Length 2129;
QY	Best Local Similarity	25.7%; Pred. No. 4,1e-81;
Db	Matches	598; Conservative 392; Mismatches 828; Indels 508; Gaps 84
QY	3 LGTRVWLRENGOHPPSTVNSCAE-----GIVFRTDYGVFT---YKOSTITTHQKV-T 52	
Db	13 IQGEVVMVKPQ-----NTTSEFAVPFGAIVRTKTEQTQLVCDNRKNQFWVPADVVK 64	
QY	53 AMHPINEGVDMASTLTHGSGIMVYNLFQYKRNQIMTYGSIILASVNPQPIAGLYEP 112	
Db	65 AMHTISQDVDEMTLGLDLEQYTIIRNIQNRPAKOLITYTSSMLVAINPYQ-ILEPTYN 123	
QY	113 ATMEQYSRRHIGELPPHIFAIANECYRCIMKKNQOCILIKGSGAGKTESTLYIKPLTS 172	
Db	124 REIQYLRKSLAEELPPHIFALSDNAPFORLQKENCQCVIISGSGAGKTESTLYIQYLA 183	
QY	173 VISOQSLELSLKEKTSQCYERALTESPIMEAFGNKATVYNNSSRFGKTVOLINCOGKN 232	
Db	184 AIS-----GKRSWIEQOILTEANPIMEAFGNKATVANDSSRFGKTIETRTQGA 234	
QY	233 QGGRIVDCILSSONRVNRONPGERNYHIFYALLAGLEHREBEFYL--STPENYHYNOS 290	
Db	235 QGARIQYLL-ERSRIVFQSRDERNYHIFYCMLAGLSTAFERRLKIQEOSPQYHNYLAG 293	
QY	291 GCVEKRTISDOESFREVVTAMDVQFSEBEVREVRLLAGILHNGNIEPITA-----GGA 345	
Db	294 GCFTLLPGRGDAKDFADIPAMKVLSFKBEWVSIISLLAAILHLGLRFTATEVANLAT 353	
QY	346 QVSFKTALGRSAELLGLDPTQLTDLATDQSMFLRGEELITPLNVQOAVDSRDSLAMALYA 405	
Db	354 BIDDPNPNQRAVQALLGIPISALNMAALTQRTIIVVHEBHTVTSLSKEAALIEGRDAFVKSLVD 413	
QY	406 CCFEYVVIKKINSIRKGNED--FKSIGILDIIPGFENPEVNVHPEQPINYANETKLOEFNNG 463	
Db	414 GIFAVIIVRINETINKQVDQPMNWSIGVLDIFEFENFPDNNFSQOLCINYNENMLQOQFVG 473	
QY	464 IFSLEQLEYSEREGVWEMEDIMWIDNECDPLI-EKKIGLALINSESHFOATDSTILLEK 522	
Db	474 IFKNEQDEYNQEHIMWQHIPEODNQIIDLIGMKRMNMLSLDSESKPFKGTQDTILLEK 533	
QY	523 HSGHANHFYKPRVAVANN-FGVKAYAGEVQYDVGILEKRNQTFPDILLNLRESRDF 581	

```

Db      534 HVOHGRRSIYVKGKTTQTSIFGIRHAYGVMMNPLFLEBNRDSFGDGRITLQORSTNXY 593
Qy      582 IYDLFPHVBSRNNOPTLKCGSKHRRPTVSQFVDSLHSIMATLSNPPFVCIKPMQ 641
Db      594 LVDFIPH---ENPMDTAK-----KOPTLCYKFR-NSLDMIMRTLSQMHFYFICIRPNEY 644
Qy      642 KMPDQDQAVLNQRLYSGMLFTRIRKQAVARRPFODIYKRYKYLMENTLALPE--DYR 699
Db      645 KKPNDKDELQVRLQYLSGMMETARIRBAGYPIRHAIVRAVERIYLLVPVGLLEGCDOR 704
Qy      700 GKSTLSLQYDASNSWQGLKTVFIRBSLEQCLEKREHVSHAMVIRAHVGLPLAK 759
Db      705 KLAQICEVALPADSDROYGKTCFLRDEDDASLEQSRQLMKSIIVTIQIRIRVLRFR 764
Qy      760 QYRKVLQCVITIQKNAFLRRRFLHKK-----AAIVQKOL----- 798
Db      765 YMKRYEALITVQRYKGRLOKRRKQYVMKQGFHRLQACIAQOULTTKFTMRCRTIKLOA 824
Qy      799 --RGQIARRVYQQLAEKEQEKEKQOE-----EEEKKKREBERERER 840
Db      825 LBRGVLYRKDFOKLERRKQNGKKEBELLKLAKEAEHLRLQOLKQKEREHQEOE 884
Qy      841 ERREBELRAQOEETKQOELEALQSKQKEAEITRELEK---OKEN----- 883
Db      885 KTLQESORLKAEMAAENALMAAVQOKRRTPVKQBPAPLQALNSLPPPTTLIVA 944
Qy      884 ----KQVEILRL-----EKEIEDLQ-R-MEQOELS-----LTEASL-- 915
Db      945 PLPTIRASAVTRNTIPEBPPTIDVSSKOMVDVFRFLDEBDVSGPLGPNKEXKSMF 1004
Qy      916 -OKLOERRDOELRLLEEACRAQOEFLSINFEIDECVY-----NIERS---LSGG 963
Db      1005 EQALRRRQVPTKLSRPA-----LRKLNINISSDITILPKSVNNIDTSDPSLTXA 1058
Qy      964 SESSSLASACEKKNFNFSPQYPRREEVDEGEADDAIKOSPNSEHGHSDQRTSGI- 1022
Db      1059 ATYFGGATAOHERKP-----LKSILKHEPIDEMASKAIIW 1095
Qy      1023 ----RTSDSSEEDPVMNDTV-VPTSPASDTYLALPSVQDSGLNNSG-SGSESTYCMQ 1076
Db      1096 LTTILRMGD-----LPDVSSPTLHFDNENLMS-----DLASLNTSDSYKPLRFLVQ 1144
Qy      1077 NAGDLPSPGDYDYPDDYEDGALITSGSS---VTSNSGY-----SQMSPDY 1120
Db      1145 SQRRIPKPLASGKEQEFYQHMLNVPITLHEKIHIIIGGIIKNSLRCPPEKEPERHL 1204
Qy      1121 KCSVGTNNSGAVRFSSEG-AQSSFEDESEDPDRFDTDELISYRDSVYSCVTLDPFHS 1179
Db      1205 R-----SFMKQGTALQATPPLQRLERTLVNGPRC--QPPSLFELHAIIRGRHP 1250
Qy      1180 PLVMKGLNMSWRKRCVYLDFTPLFRSKQELAKO-----GWLH 1219
Db      1251 -RLDLHMLDGOQRRLQVDAAST-----AREVNVNLCQSMGLTDTFGGLVMSLNGKLM 1303
Qy      1220 KKGGS-----STLSRRYMKRW--FVLROSKLAFVENDSEBK-----LK- 1257
Db      1304 PLGAGEHVLDAISECEQOLARPMLYIRKEMFATWYDSNDPKATOLYKQILNGLKC 1363
Qy      1258 GIVEVNTAKEIIDNTT---KENGIDIMADRTPHLIAESPBD-----ASQWFSVLS 1305
Db      1364 GEYCRSEBKDIAMVNCALACFEVYGPGEILRLKSEITAFVPSDLAPGRATENWRLIA 1423
Qy      1306 QVHASTDEIQEHMDQANPQNAVGLDGLIDVCAQSDSPDR---NSFVITIANRVL 1361
Db      1424 ATYKSSYKVEENDLLLEAQKAK-----EDICLFAILSWPMHSHLFEVV----- 1470
Qy      1362 HCNADTPEEMHMTITLQSKSDTRVEGOFIYRWGLHKE-----VGNSPKXMSLKLTKR 1416
Db      1471 --RKEGPKQSDMLMGINSAGFLIDETQVLAASCFBSVLKVNYESDDKLHMTFQYV 1528
Qy      1417 WPLYLINS-----LDYKSSSEKNAKLKLTGLVLSL--CSVVPDEKIPKETGYW 1463

```

```

Db      1529 NEVLQSSAODANEVITNMLDNLQRSSYGVALDPVEGDLDELCLVINGDLIFEBAGVT 1588
Qy      1464 NNTVY--GRKCYRLUTKLINEATRWSSVIGVNTDTKAPIDPTQOOLQIDIKENCLN--- 1518
Db      1589 GAOIAGNAQDCYR---GCYN--GOMGQFLAGNVAVLTLTPRSKTLQDILABGRQOEP 1643
Qy      1519 -----SDVEQIYK-----RNPILRYTHNPLHSPRLPLPYGIDINLNL 1556
Db      1644 KPTPRANYSRRQHNISQLAESHPRELDSDKAPLSKFSPEPLKAPIL----- 1691
Qy      1557 KDKGYTLDEDAIKIFNSIQ-----LEMSD-----PIP---IIQGLQCHDR 1599
Db      1692 -----KAVKXPPLEQOALVMEHNLILKTYGDIARSNLVNTDLIFQPLQ--HPL- 1739
Qy      1600 PLRDELQOLIKQTKNVPKGVSNLYSQILTSCFLPS-----RGLIKLYKFN----- 1651
Db      1740 -LCDELYCOLMKQLS--DNPSSESEKRGNDLYLATGLVAPSVLVKRELIITLIRRADYL 1796
Qy      1652 ----LKRIEOPPGTEMEKYA-LFTYESLKKTKC-----REPVSARDEIBALIHQEMTS 1701
Db      1797 ADACLKRLKRSIAGQORKKAPHLIEVGIQ--RCHLYHKIYFPD-DVTEA----- 1845
Qy      1702 TYVCGGSCCKTINSHTTAGVEVKLIRGLAMEDSRMNFALFYNGHYDKAIESRTVYA 1761
Db      1846 -----PEISHTRGAEIADIQRLIELK-SPVQYSIFLTKGVRVYAMPBEEFVP 1893
Qy      1762 DYIAK----FEKLAATSEVGDLPWKFPYFLYCFLTDVNPXSV--KFAFMPEQAHAVI 1815
Db      1894 DFITQILYLRQORTIRISDQYQHFMRKML--NNHGEDLNGDMIFSYPOELHKKYL 1951
Qy      1816 HGHHPAREBNLOVALRALQYLOQDYTLHAIPLREEVYSLORLKARISQSTKTFTPCR 1875
Db      1952 KGYVPIDEQASRLAIL-----YASADHVSILQRLP----- 1982
Qy      1876 LKRRTSFLBGLRFSRTGSSVYRQKVEEOMLDMIKESVSASAASITDKRKQCGANQ 1935
Db      1983 -----EVLTRLIP-----DLPLQTVAMERQQLIPKYR--DHLE 2017
Qy      1936 EOMAKYMALIKEMPGYSTLPDV--EKEGCGPOELMGVSADAVSVYKRGSGRPLEVQ 1994
Db      2018 DHAKLILFOELSHFCFSGTFVVKQONDALPETLLAINSTGFHMLDPTTKELIRSYE 2077
Qy      1995 YEHILSFGAPLANTYKI-----VDERELLFETSEVVDVAKMKAYI 2036
Db      2078 YGQ-LGIVSSGKNHPIIRGNMIGASKLSCSTGOYKQMDLLASVY 2122

RESULT 10
Q9NH54
ID Q9NH54; PRELIMINARY; PRT; 2121 AA.
AC Q9NH54;
DT 01-OCT-2000 (TRENBLREL. 15, Created)
DT 01-OCT-2000 (TRENBLREL. 15, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Myosin VII-like protein.
GN MYO28B1 OR CG6976.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RA Anderson U.B., Yamashita R.A., Sellers J.R.;
RT "Complete cDNA for an unconventional Myosin (Class VII) in Drosophila
  melanogaster."
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF233269; AAF34810.1; -.
DR HSSP; P08799; 1MD.
DR FlyBase; FBgn0040299; Myo28B1.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.

```


QY	1663	EMEXKA-LFTYESLKTGTC-----REFPSPDELEALHNRQEMSTVYCHGGSGOKITN	1716
Db	1804	ORKKAFHLEIWEVGIOQ-RCLHIIYHKTYPD-DTEA-----FEIE	1841
QY	1717	SHTTAGEVVEKLIIRGLAMESRNMALFEYNGHVDKAIESTRTVADVAK----	FEKLLA 1772
Db	1842	SHTRGAELIADIAGLBLEK-SPVGSITLKTGDRVYANPEEPEDFTTOLLVYMRQORT	1900
QY	1773	TSEVGGLFPMKTFPKLYCFPLTDNVPKDSV--EPAFMEQAHNAVHGHNAPEENLOYLA	1830
Db	1901	IRSIIDGQYQHFMKML--NNHGEGLNGDMITPSYPOELHKYLGYYPIDCEQASRLA	1958
QY	1831	ALRLQYLOGDVTLLHAIIPLEBEVYSLOQLKARISGSTFTFPCERLEKRTSFLBGTUUR	1850
Db	1959	IL-----VYSADHVSLOLRIP-----	1974
QY	1891	SFRTSVVRKQVEEQOMLDMWIKKEVSSARASIIDKMRFGOMNOEQAMAKYMLIKWP	1950
Db	1975	-----EVLTRLRPE---DLIRPLQYVAEMRQOILPKYR-DHLYEDHAKILFLOELSHFA	2024
QY	1951	GYGSLTLPDV-ECKEGSGPOELMWLVSAVSAVYKRGEPLEVEFOYEHILSFGAPLANTY	2009
Db	2025	CFGSTFFVVKQONDALPETLIIAINSTGFHMLDPTTKELIRSYEQ-LGIWMSGKNHF	2083
QY	2010	KI-----VYDERELLEFTSEVDVDAKLMAYI	2036
Db	2084	HIRFGNMIGASQLCSTTGYGMDDLLASV	2114

RESULT	11		
ID	Q99M26	PRELIMINARY,	PRT, 2113 AA.
AC	Q99M26,		
DT	01-JUN-2001	(TEMBLrel. 17, Created)	
DT	01-JUN-2001	(TEMBLrel. 17, Last sequence update)	
DT	01-JUN-2002	(TEMBLrel. 21, Last annotation update)	
DE	Myosin-VIib.		
CN	MYO7B.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euteleota; Rodentia; Sciurognathi; Muridae; Murinae; Mus		
OX	NCBI_TaxID=10090;		
RA	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=21295050; PubMed=11401444;		
RA	Chen Z.-Y., Haasson T., Zhang D.-S., Schender B.J., Dertler B.H.,		
RA	Moosker M.S., Corey D.P.;		
RT	"Myosin VIb, a novel unconventional myosin, is a constituent of		
RT	microvilli in transporting epithelia."		
RL	Genomics 72:285-296(2001).		
CC	-1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.		
DR	EMBL, AF242411; AAC28339.1; -.		
DR	HSSP; P08799; 1MND.		
DR	MGD; MGI:107709; Myo7b.		
DR	InterPro; IPR002106; AACRNA_11gase11.		
DR	InterPro; IPR000299; Band 4.1.		
DR	InterPro; IPR000871; Beta_lactamase_A.		
DR	InterPro; IPR000048; IQ region.		
DR	InterPro; IPR001609; myosin_head.		
DR	InterPro; IPR000857; MYTH4.		
DR	InterPro; IPR001452; SH3.		
DR	InterPro; IPR000508; SigPtaase.		
DR	Pfam; PF00612; IQ; 5.		
DR	Pfam; PF00063; myosin_head; 1.		
DR	Pfam; PF00784; MYTH4; 2.		
DR	Pfam; PF00018; SH3; 1.		
DR	PRINTS; PR00193; MYOSINHEAVY.		
DR	ProDom; PD000355; myosin_head; 1.		
DR	SMART; SM00295; B41; 2.		
DR	SMART; SM00015; IQ; 4.		
DR	SMART; SM00242; MYSC; 1.		
DR	SMART; SM00139; MYTH4; 2.		
DR	SMART; SM00136; SH3; 1.		

DR PROSITE: AA_TRNA_LIGASE_II2; UNKNOWN_1.
DR PROSITE: P500357; BAND_41_3; 2
DR PROSITE: P500146; BETA_LACTAMASE_A; UNKNOWN_1.
DR PROSITE: P55002; SH3_1.
DR PROSITE: P500761; SPASE_1_3; UNKNOWN_1.
KW SH3 domain.
SQ SEQUENCE: 213 AA; 240785 MW; FC8B09163DD0C869 CRC64;

Query Match	15.2%;	Score 1635;	DB 11;	Length 2113;
Best Local Similarity	24.7%;	Pred. No. 2e-77;		
Matches 577;	Conservative 385;	Mismatches 830;	Indels 542;	Gaps 82

```

0Y 5 GTRWLRNGCHFSTYNSCAELGVFRTDYGQVFTYQSNITTH-----QKXAMPPT 57
Db 7 GDHWLDPDS-----SSKTGVAIGGIIVKXTKGKTLIEDDEGEKHHVAHJADSLTLEPMHPN 62
0Y 58 NEEGVDDMA$TLELHGGSIMVNLFORYKRNQIMWTYIGSILASVNPYCPJAGIYEBATBQO 117
Db 63 SAQGVDDMIRLGDINEAGVANNLIRYQHKITYTGSIIIVAVNPFQULP-LYTEQVQI 121
0Y 118 YSRHRLGELPHIFAIANECYRCILMKRHDNOCILKGBSGAKTESTKLILKPLSVISQO 177
Db 122 YSRHMGELPHIFAIANSCYFNKKKNRQDCCISGSGAKTETTKLILOPLATVISOQ 181
0Y 178 SLELSKERTSCVERALIESSPIMEAFGNATVYNNSSRPKFPQOLNICQKNIQGGRI 237
Db 182 H-----SWIEQVLEANPILPEAFGNAKTIRINDSSRPGKYDIHFHNSG-VLAGAS 231
0Y 238 VDCILSSGNRYVRNPGERNHIFALLAGLEHERREFFYLTSTPNYHYLQNSGVEDBT 297
Db 232 IEHFLLESRCROKPEERNYHIFCMLGMSPEBKQULSAMBETHILTMGSCITSSBG 291
0Y 298 ISDQSFREYVTAMDVMOFSKEBEVREVSRLAGILHLNINIEPITA-----GGAQVSFKTA 352
Db 292 LSDAKDVHVASAKKILQFSDSEMMDISKLALAILHGNVGMAVFNELSDSDVMEFPA 351
0Y 353 LGR$AELLGLDPTOLTDLTPRSMFLRGEELTPLNVQOAVDSRSLMALYACCFENVY 412
Db 352 PPLAMKILEVOHQAIRDCILIKHTTIPVLEPF$RPNVNIQATDRDAPFVIGYIGRFLQWIV 411
0Y 413 KKINSRI-----KQNEPFK$IGLIDIGFENFEVNHQFNINAVNEKLO$YFNKHIFS 466
Db 412 KKINAIFITPOADPQVNRRAIGLLDIGFENFQNNSEOLCINPANEHLQOFPVGHVPT 471
0Y 467 LEOLEY$REGVWEDIDWINGECJDLLE-KKLGALLINE$HPQATDSTLLEKJHSQ 525
Db 472 MEQSEYLESENITWNYIHTDNOPIIDMALPKMSIISLDEBSRFPQSTDVMTLOKANSI 531
0Y 526 HANNHFYKPR-VAVNPNFVGKIVAGEVOYDV$GILEKURDTPFRDDLMLLRESRPDIYD 584
Db 532 HANNK$PLSPRSIHOTRFGIAHPGADVYYQAGEFLKXRDVLTSDIILLHNSKOKFLKE 591
0Y 595 LFEHVSFN-----NODTLKCG$KRRP-TV$QPKVDLSH$LAUTLESSNPF 631
Db 592 IFNVDS$QTKLGHGTICQVKAG$OLF$KSDSIKRPVTLASQFK-OSLQOLMKRILTNCPY 650
0Y 632 FVRCIKPMQKMPDQFDQAVVUNLQ$YSGMLETYVIRKAGAVRRPDPQYKRYKYM-- 689
Db 631 FVRCIKPEYK$PLFDRELICQ$IRYS$GMEYAIR$SGP$IRITTFD$ESORF$RL$S 710
0Y 690 -RNLLP$EDV$GKTSLLQYDAS$ENQLO$GKTVKFLRESLEQKLEKREBEVSHAAM-- 746
Db 711 PERMQFQKPRQMTLHIDLCLGTDKMKVGT$KFLMDHODTVLEIR$SO$LDGA$IRI 770
0Y 747 --VIRAH-----VIGF-----LARKOYR 762
Db 771 QVLRGHKYRKEFLQR$RAAVTLQ$GMRGYSQ$RK$FKLLVGFERLQ$AIR$HLLMRQO 830
0Y 763 KVL$YCVYLIQK$YR$FLLRR$F$HLK$K$A$YI$FQ$OL$BO$QI$AR$V$RO-----L$ARK 814
Db 831 AMRORIVOLQ$R$CR$Y$IV$QO$Q$AK$R$R$V$VIIQ$AH$R$M$V$VR$K$S$TWQ$K$ST$P$OV$IT$A-- 888
0Y 815 REQ$EKKQ$E$E$E$K$G$R$E$E$R$E$R$E$R$E-----AELPAQ$E$E$TRKQ$O$ELALQ$S 867

```



```

Db      889  --KEPAAOVANHERKRSIYDVTDTAMVEKYVGLPAMIGCGECPAPTRFEDLEV--KT 944
Qy      868  OKAEELTRELEKKEKKNQVBEILRLKEKIEDLORMKEQOELSTE-----A 913
Db      945  OKKHE-----VDDTVPMAMPEEVDLSAETFPKFAVTVQKA 985
Qy      914  SLOKORRDEQLRLSEECACRAA--QEFLESINF--DEIDECVRNERSLSGSEFSSEL 970
Db      986  SHHIIQPKLKYPLLYHENDDHSAIDWVILIRFNGDLPEPVYVGRNSLTGSSVWR---- 1042
Qy      971  ABAACEKKNFNFSQRYPEBEVDGEBADDAFKDSFNPSSEKSHSQORTSGITSDSSE 1030
Db      1043  -----QIHDKLGKDS--VTQHRSSQOVASQLNFGGEARK 1074
Qy      1031  EDPYMDTVVPSPASDSTV---LLAPVOD-----SGLSNLSGSESTYCMQPN 1077
Db      1075  FDEPISDRPMNLKFNHFIYGVAIMRPLGDLDEIYCOICQKLSNNTSSRPMWILLSLC 1134
Qy      1078  AGDLPSPPDGDYDDODDYEDGATSGS-----SVFSNSYCSQ--W----- 1116
Db      1135  LGGPPSEREMKTLNLPISGPPSYGPFCAERLQRTFANGVRABPPTWLEQAVSKKH 1194
Qy      1117  -----SPDYRCVGTYNSSGAYRFSSEGAQS-----SFEDEEDPDSRPDTDELSY 1163
Db      1195  PIQVILATGRSLTISVDASASTSREICQHVAKOGKGLDNLGFSLQVAVVDKFMSLGSGCDH 1254
Qy      1164  RRDVSVCYTLPLPFHSLFKKGLNMSKMRKCVLDEFTL--WFRKQELMKGMHKKXG 1222
Db      1255  LMDAVACQEQOLABERG---ESORQAPWR---IYFKFEFTPMHDSQEDVSTELIYHO- 1306
Qy      1223  GGSSTSRNMKRWFLRQSKLMPENDSEELKGTVEV--RTAKEIINDTKENGIDI 1280
Db      1307  -----VLRGVMSGEVNFKEBEL---VELLHNHCYVQIGATYKSAVQOE 1347
Qy      1281  IMAD-----RTFHILASPEBASQWFSVLSOVNA-----STDQEIQENHDAQNP 1325
Db      1348  LLPSCVPSKLYRT---KSPE---KMASLVTAAHAKAOTQSKATPLAREQVTEAA-- 1397
Qy      1326  QNNAVGTLDVGLISVCASDSPDRPNPSFVITANRVLHCNADTPEBNNHTTLLQRSKGT 1385
Db      1398  -RLMLPLLSRLREVTTLTSGFRLPKTQVLAIN-----WKGMFLDQKER 1441
Qy      1386  RVEGOEFT--VRGWL--HKEVKNSPKMSLSKLKRWFLVTHNSLDYKSSSEKNAKLGTVLV 1443
Db      1442  TLLGLSPAEWGVAVNRDAGKKLLATLQEEYEFVSPSY--ALAEVVALFLGLKHE 1498
Qy      1444  NSLCSVVPPEKIFKKTGTWNTVYG-RKHCYRLYTK--LL-----NEATRWSSV 1490
Db      1499  RSVFAMALQDRATDD-----ITLPPKKGDLLILTKKQGLLASEMWALQONDRGTGTL 1553
Qy      1491  IQWVDTKAP-IDTPQQLI-----ODIK--ENCLNSDVVEQIY----- 1526
Db      1554  VPTACLYTTPSVTKPSTQLSLSLAMSPBEKKLAAQEVRLPELPDQLTESPTLEEFST 1613
Qy      1527  -----KRNPIRLYTHNPLHSPLLPLPYGDINILNKLKQGYTTLQD 1566
Db      1614  QPRAPEKETISRRAAMPARSRHLMAYSPERLQPL-----LKSHYDK--AKLND 1662
Qy      1567  BAIRIF-----NSLOQLBSM-----SDPIPIIGLITQGHDLRLPLRDE 1605
Db      1663  AACQIFALILKYGDYPSRQSWHSLLELTDQMFSLALQDP-----ALQDEL 1707
Qy      1606  YCOLIKOT--NKYPHGSVGNLYSKQILTCLSTPLPSRQILKYKFLKRLIREQPPGE 1663
Db      1708  YCOLIKQLTNNSI---RFESEERAWQMLMLCTGLPFGKTLPLH-----AQKIDSR 1755
Qy      1664  MEK-VALLFTYESLK---KTKCREFPVSRDEIBAL-----IHRQEMTSYVYCHGGSCKI 1713
Db      1756  KKRPLALDCSRRLHRYLVAVGPRKQPHNDVEKAAEQNVSKLHNE-----VYLPNDISKM 1810
Qy      1714  TINSHTAGEVVKLRIG---LAMEDSRMFFALFEYNHGVDAI-----ESRTVY 1760

```

```

Db      1811  EYSSSSRVRDLCEGI--GTRQLASMDGCSLEFIK-----TDKVISLKGDFPFDSLROV 1863
Qy      1764  ADVLAKFEKALATSEVGLPMKVFYKLYCFGLDTONVPKDSVEFAPFEQABAVIHCHRP 1820
Db      1864  SDWYKKNRPQKEGAGAV--TLPIYVFFMKRLMNLV--TPKQDN-----ADTIHAYOE 1912
Qy      1821  ABEENIQLVLAIRLOYLGQDYTLHAAPLPEBEVSLORLAKRISOSTYTFPCRLEKR 1880
Db      1913  LE-----KYLKG-----FHKC-----GRE 1926
Qy      1881  TSFLEGLTRSPFRSTSVVRQKVEEOMLDMWIKREVSSARASIIDKMRF-----QG 1932
Db      1927  DAHLGGLICKIQFGSDSSQLASVSKVLKEIWPQNLTRLMS--BEWKSILLBEDCKNR 1984
Qy      1933  MNQEQAMAKYVALIKEMPGYSTLPDV--ECKEGFPQOELMVLGSADAVSVYKRGGRPLE 1991
Db      1985  KTVABKAYEFLKMYMRWPFSGAFPEVQOTSPSPDILLAINHGLVLIHPKKEILN 2044
Qy      1992  VFQYHILSPGAPLANTYKIVV---DERELIFTSEVVDYAKLMKAYISMIV 2040
Db      2045  TYPFTKISSWS--GNTYFHMLGSLGQSGRLLCETSLGYKMDLLTSYVQOLL 2096

RESULT 12
Q9XE14
AC Q9XE14 PRELIMINARY; PRT; 1529 AA.
ID Q9XE14;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Unconventional myosin heavy chain.
GN MYO1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
OC Panicoideae; Andropogoneae; Zea.
OX NCBI_TaxId=4577;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu L., Pegacreta T.C.;
RT "Myosin gene family of maize.";
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF104924; AAD17931.2; -.
DR HSSP; P08799; 1MND.
DR InterPro; IPR002710; DIL.
DR InterPro; IPR000048; IQ_region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR004009; myosin_N.
DR Pfam; PF01843; DIL; 1.
DR Pfam; PF00612; IQ; 6.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR ProDom; PD003376; DIL; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PSS0096; IQ; 1.
SQ SEQUENCE 1529 AA; 173455 MW; 22BE5947DC3EB3D4 CRC64;

Query Match 14.1%; Score 1512.5; DB 10; Length 1529;
Best Local Similarity 34.6%; Pred. No. 3.7e-71;
Matches 387; Conservative 187; Mismatches 403; Indels 143; Gaps 26;

Qy 5 GTRVNLRE--NGCHPSTVNSCAEGLIVERTDYGQVFTYKQSTITTHQKTYAHNPTNEGVD 63
Db 10 GSHVWEDPNLAWIDGEVVISIKNNEVHVQTSNGKVTYDRSKVFPKDEA--PFG--GVD 65
Qy 64 DNASITLHGSIMYNLFCORYKRNQIMWTYIGSILASVNPYOPDIAGIYEPATMEYSRRL 123
Db 66 DMTRLSYLHEPVLQNLATRIYELNELYTTSILAVNPFQVLPHLIDTHMMEQKGADEF 125
Qy 124 GELPHIFAIANECYRCIMKRHDNOCILIKGSGAGKTESTKLILKFLSVISQOGLSL 183

```



```

Db      133 HPOTFALQPEBGSRLARQDLSSGVEDMTLLDLHNSALMLNRLRYKGLITVFASSIL 192
Qy      98 ASVNPYOPAGIAYEPATMEQYSRRHGLPPIHFAIANECYRCMKRMDQCILIGESG 157
Db      193 IANPYMPFADAYGLVAKQYAGRPGLSLPPHLPALCAAAHAAV--PSQVAVVIGESG 249
Qy      158 AGTSTKTLILKFLSVISQOSELSELEKTSCEVERALIESSPIMEAFGNKATVYNNSSR 217
Db      250 SGTSTKTLVMOYLAAVPPGGGAS-----AVITEQLLEAALPILFANGNARTARNDNSSR 304
Qy      218 EGFVQVQINOCQKNIQGRIVDCILSQNVKQNPBERNYHIFYALLAGLEHEEBEF 277
Db      305 FGRLVEVYF-KSAIVAKITQYLL-EKSRIVTQAPERNYHVFYELGLSESTERSKYG 362
Qy      278 LSTPENHYALNQ--SGCVEDKTSIDQSPREVIITAMVMOFSKBEYEVSRLLAGILHLG 335
Db      363 LLEADKFLYNQCATDCASGRV--DMESLO--GANOVLGVSGERBGIVLAAVHLHG 417
Qy      336 NIEF---ITAG--GAQVSFKTAGRSALLGLDPTQLTALTRQSMFLNGEELTPPLV 389
Db      418 NVYFHRQLRHQGEVGVSGDAEIKMAHLHLISADLHRALTSRTTEARAEHLHTPLGI 477
Qy      390 QQAVDSDSLAMALYACCFEYVVKIKINSRK--GNDFKSGIGLIDFGFENFEVNHPEQ 447
Db      478 DQALDADAFKALYAGLFNWLVSRLNSIVQKGTDAHRISILDFGFEDELAENSPFOL 537
Qy      448 NIVANENKLOEYFNKHFISLEOLEYSEGLVMDIDWIDGECGLDIEKK-LGLALLINE 506
Db      538 CINYANENLOLYNKVHFKLEQAEYARELERLEPTLAMDYLPYIHLAKKPVGICHLDD 597
Qy      507 ESHFPOATDSTLLEKHSQHANHFYVPRVAVNNGVKNYAGEVOYVGLIEKORDTE 566
Db      598 ESNFPRATDLSFLEKCHYNHALSELVAPRIGAEFGVTHYAGQVWVCVDFDKORDAL 657
Qy      567 RDLLMLNRSRDEFTYDLFEH--SSKNMODTLKCS-----KRRPVSQGFKVDL 618
Db      658 RGVLELASSRPLVGLLTKLRAQDACKTLPKSGNGRVTMKPTTPVAARF-ADSL 716
Qy      619 HSLMATLSSNPFVRCIKENMKMPQPOQAVLNLQRYSGMLETVRIKAGVYARRPF 678
Db      717 QQLQSGRGCHPWFVRCIKENOKHALRMDPCVLOQLRYLGMIDTIQIRQRYPRALRF 776
Qy      679 QDFYKRYKVLNRN-LALPEDVRKCTSLQLYDASNSE---WQDKTYVFLRESLEQLE 734
Db      777 QHVERRYRHLPSPLARGTYPRELCPALREBAMPRTGVEGPDYQLGATRVFLREALRALE 836
Qy      735 KRBEESHAAMVIRAHVIGFLARKQYKLYCVIIQKRYRAFLLRRPLHKKXAIYF 794
Db      837 SGRTERLRRAVASVQRHVHGMVLRQLARRQAAATRLQARWQORAQOQRYERLRKALTA 896
Qy      795 QKOLRGQIARRVYRQLAEKREOEKKEEKKKEEEREREREREREREAELRAQOEE 854
Db      897 QRLMRGQARRRRQQLRSDHRRQEAR-----BAQRAREARA----- 935
Qy      855 TRKOQLEALQKSK--AEL 873
Db      936 --KQAVLERSQLSYLDIPAL 954

```

RESULT 14

```

Q9UNJ2  PRELIMINARY; PRT; 2548 AA.
ID      Q9UNJ2
AC      01-MAY-2000 (TEMBLrel. 13, Created)
DT      01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT      01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE      Myosin-IXa.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.

```

```

RX      MEDLINE=99339979; PubMed=10409426;
RA      Gorman S.W., Haider N.B., Grieshammer U., Swiderski R.E., Kim E.,
RA      Welch J.W., Seabry C., Leng S.R., Carmi R., Sheffield V.C., Dull D.M.;
RT      "The cloning and developmental expression of unconventional myosin IXA
RT      (MYO9A) a gene in the barbed-biased syndrome (BBS4) region at
RT      chromosome 15q22-q23."
RL      Genomics 59:150-160(1999).
RL      EMBL; AF117886; AAD49195.1; -.
DR      HSSP; P08799; 1MDN.
DR      InterPro; IPR002219; DAG-pe-bind.
DR      InterPro; IPR000048; IQ_region.
DR      InterPro; IPR001609; myosin_head.
DR      InterPro; IPR000159; RA_domain.
DR      InterPro; IPR000198; RhogAP.
DR      Pfam; PF00130; DAG-pe-bind; 1.
DR      Pfam; PF00612; IQ_5.
DR      Pfam; PF00063; myosin_head; 2.
DR      Pfam; PF00788; RA; 1.
DR      Pfam; PF00620; RhogAP; 1.
DR      PRINTS; PR00193; MYOSINHAIV.
DR      ProDom; PD000355; myosin_head; 1.
DR      SMART; SM00109; C1; 1.
DR      SMART; SM00015; IQ_5.
DR      SMART; SM00242; MISC; 1.
DR      SMART; SM00314; RA; 1.
DR      SMART; SM00324; RhogAP; 1.
DR      SMART; PS00479; DAG-pe-bind-dom_1; UNKNOWN_1.
DR      PROSITE; PS50081; DAG-pe-bind-dom_2; 1.
DR      PROSITE; PS50096; IQ_1-bind-dom_2; 1.
SQ      SEQUENCE 2548 AA; 292703 MW; B93B76C2A0E9A356 CRC64;

```

Query Match 13.9%; Score 1492; DB 4; Length 2548;

Best Local Similarity 24.3%; Pred. No. 8.8e-70;

Matches 541; Conservative 349; Mismatches 674; Indels 660; Gaps 79;

```

Qy      35 YGVFTYKQSTTTHQKVT--MHPTNEGVDMASLTFLKSGSLMTNLFQRYKKNQWT 91
Db      118 YGSLQSWLRVTEERRRMERGFLLPOQKDFDLCSLDLEKTLLENLRDFKHEKIYT 177
Qy      92 YIGSLIASVNPQPIAGIAYEPATMEQYSRRHGLPPIHFAIANECYRCMKRMDQCILIG 151
Db      178 YVGSILIVINPK-FLPIYNPKYVMYDNDHQLKPEPHYAVADVAHYAMLORKKNOCTV 236
Qy      152 IKGESGAGTSESTKILKFLSVISQOSELSELEKTSCEVERALIESSPIMEAFGNKATVY 211
Db      237 ISESGSGKTQSTNFIHLHLSQGF-----ASGVEQLITAGPLAEFGNAKTAH 289
Qy      212 NNNSSRFGKFGVQVLANICQKNIQGRIVDCILSQNVKQNPBERNYHIFYALLAGLEHE 271
Db      290 NNNSSRFGKFGVQVLANICQKNIQGRIVDCILSQNVKQNPBERNYHIFYALLAGLEHE 348
Qy      272 EREEFYSTPENHYALNQ-----SGCVEDK-----TISDQ-----SPREVIITAMV 313
Db      349 ERSAPHLKQPEEYHNLNQITTKKPLRQSDMDYTYDSEPDCTFEGEDLRHDFRLQLANEM 408
Qy      314 MFSKEBEVREVSRLLAGILHLGNI-----EFTTAGAQVSFKTAGRSALLGLDPTQLT 369
Db      409 VGFLPKTRQIFSLLSAILHLGNI--CYKKKTYRDSIDICNPEVLPVLSLEVEKEMLPFE 468
Qy      370 ALTVKRYTVVGEKTLIPYKLAELAVTVRNSMASLSALFDMVIFRINHALNSKDLHNHT 528
Db      427 ---STGILIDFGFENFEVNHPEQFNIYANENKLOEYFNKHFISLEOLEYSEGLVMDID 483
Qy      529 KTLSTGVLIDIFEFEEYENNSPEQFCINPNERLQYFNOHIFKLQOEBEYRTGISMHNID 588
Db      484 WIDGECGLDIEKK-LGLALLINESHPQATDSTLLEKHSQHANHFYVPRVAVVNF 542
Qy      589 YIDNVCINLISKRTGLHLHLDSESNPQATNQLDLFKKHQHDNSYIEPRAMPRF 648
Qy      543 GVKHYAGEVOYVNRGILEKRDTPFDLLNLRESRFDFTYDL----- 585

```

Db 649 IIKHVAQKVKYKQFREQKTDHMPDIYALLRSSKNAPISGMIGIDPAVFRMALRAF 708
 QY 586 -----FEHVSRRN-----NOTTLKCG----- 601
 Db 709 PRAMVAFRAGKRNIRKTKGHDTAPCALIKSMDSPFIQHPHQBSLEILQRCKEKXS 768
 QY 602 ----- 601
 Db 769 ITRKNPRTPLSDLOGMNALEKNQHDTPDIAMNGRTIGROSRLSGSTSLDKGIFANST 828
 QY 602 -----SKHRP 607
 Db 829 SSKTLERAGILTRKNPFKSPALPHLEVNSLKALITRLTLODRITKSLHLHKKKRP 888
 QY 608 TVSQPKVDLSLMLATLSSNPFPRCTIKPMQKAPDQAVVLNQLRYSGLMETVRI 667
 Db 889 SISAGQA-SLGLMETLGOAEFYKICRSNNEKPLRFSDDLVRQRLRYGMLETTQI 947
 QY 668 RKAQIAVRPPFODPYKRYVILMRNLALPEDVNGKCTSLQLYDASNEMQLGKTVFLRE 727
 Db 948 ROSGSSKYSFODFVSHFVILPRNIIIPS--KENIQDFPKINLNPDNYQVGKTMVFLKE 1005
 QY 728 SLEQLKEREKEV-----SHAAVYIRAHVGLAKQYRK- 763
 Db 1006 QSRQHLQDILHQBVLRIILLORFVULLCROHFLHLROASVILQRFWENYINLQKVRDA 1065
 QY 764 -----VLQCVIILQKNVRAFLRRRFLHLKKAIVFOKOLRGQIAR----- 805
 Db 1066 AVOKDAFWASAAALLQASWRAHLEKORILELRAAIVIQAKRDYRRRRHAAICTQAR 1125
 QY 806 --VYRQLAEKREOEKKKQOEBEKKKREBEREREREREAELRAQOE-----ETRK 857
 Db 1126 WKAYRE--SKRYOEORKKIILLOSTCRGFRAORFKALKEQGLRETKREVGLVNIKGYG 1182
 QY 858 QOELELQSQ--KEALITRELEKQENKQVEILREKEITEDQPKKE-----QOELSLTE 912
 Db 1183 SLEIQSDSPSEWDCSPDNRIKAECKSVITESNRISRE--SSVDCCKEPMKQOEBAQSQ 1241
 QY 913 ASLOKLQF-----RQDELRLEEBACRAOEFLSLNFDZIDEVYRIE-----RSLSCG 963
 Db 1242 SGVD-LQEDVLRERPRSLIEDLHQKVGRAKR--ESRRRZELQALFSLLELKNLSLGI 1298
 QY 964 S-----BFSSELAESACER--KPNFNFSQ-----PYDEE-----VDEGEFADDD 1001
 Db 1299 SPSEDRRWSTELVPEGLQSPRGTPDESQSLLELSEYEQSKLESYISDEG-----D 1353
 QY 1002 AKDSDNPBEHGHSDORTSGIRTSDDSR-----EDYVNDIVYPTSPSADSTVLAPSVQ 1057
 Db 1354 LQPPSPKISSSPKFRDNLASASNETSSAHLKQDTMKVVCSSSES-----ITCKPQK 1409
 QY 1058 DSGSLHNSSSGSEIYCMPONAGDLPSPDGDYDVEDDAITSGSVTFPSNYSQSKS 1117
 Db 1410 DS---FISNLSLPRFFIYPOODPLKINSQUDTISQKRLKENEDTAEALILDIRKTR-- 1464
 QY 1118 PYRCS-----VGTYNSSGAYRPSSEGAOSSPEDESDEDP:RREDTDEL----- 1161
 Db 1465 -RYHSGKQOIVPSLNTESNPVLKYLEKLNTEKEROQOQOQNEKEMEQIROOTDIL 1523
 QY 1162 -----SYRDSVYSCVTLPTFHSFLYMKGGLANSMKRNCVLTQDETFMFRSKQF-- 1211
 Db 1524 EKERKAKTIEKRIQECILVAP-----SSYOSKQVERPSSILLSTINSMKE 1570
 QY 1212 -----ALKQGLHKKGGSSSTLSRNNKGMFWLROSKLWFEJNSEKLGTV---- 1260
 Db 1571 LNVLSLSLKDALLAQKSSSAHLPPKD-----RNVYVFEKSGSPCOSSTYKELS 1621
 QY 1261 -----EYRTAKEIIDN-----TTKEN-GIDI 1280
 Db 1622 KTDHMGTOJLVACKLSNNRISKREHPRPQSYSHNSDDLREHGNARPIFTPDNNISIL 1681
 QY 1281 IMADRTFHLIABSP-----EDASQW--FSLVSQVHASTDQF:QAMDEQANPONAVCTLDV 1334
 Db 1682 VSKEA--LNSKNPQLHKEDEPAMKPVKLAGPGQRETSPRPSSV--DEQAKLHK--TWSQ 1734

QY 1335 GLIDVSC-----ASDPDRPNSFVITITANRVLHCNADTPEENHHWITLLQRSKGDTRV-- 1388
 Db 1735 GEITKLAVRQKASBDBIRPQ-----AKORFMAKQSGEKKTTRVAPT 1777
 QY 1389 -----GOEFT-----VRGMLHKVQNSPKMSLSLKLKRW 1417
 Db 1778 TQSEVSPFAGTDVPAHQFDELAAYHPTPPLSPBLPGSCRKKEKKEKPS-- 1830
 QY 1418 FVLTHNSLDYKSSSKNALKGLTULVNLGCVVPDEKIFRETGYMNTVYGRKHCVRLY 1477
 Db 1831 -----KAKRKSVKISNVALDSM-----HWQ----- 1851
 QY 1478 TKLNEATRWSSVIONVDTAPIDTPQOLIODI--KENCINSDVVEQIYKRN--PILRYT 1535
 Db 1852 -----NDSVQ-----LIASVEDKS--HDEFLLKVVNDLNDSDSKDLDVDPFKALKEPQN 1903
 QY 1536 HHPLHPLPLPYGDINLNLKDKGYTTLQDBAIFNSLQOLESMSD--PIPIIGILOT 1594
 Db 1904 IFSPYSALAMDQ-----KSIRYKDLVALFQOILEKTRLEORDSLIGSPYRV-----W 1953
 QY 1595 GHDLARPLDELQCOLIKOTN-----KVP-----HPSVGNLYSQILT- 1632
 Db 1954 VNTFVFLDB--YMEFKTSDCTATKVPTEKKRRKKSTDLVEBHNGHIFPATQYSIPTY 2012
 QY 1633 CLSCTFLPSRGLIKYLKRIEQLREOPGTEMEKYLFTYESLTKT--XC-----R 1681
 Db 2013 CEYCS-----SLWINDRASVCLCKACHKKCKLKTATACSKKYDELSR 2059
 QY 1682 EF-----VPSRDE-----IEALHROEM-----TSTVYCHGGGSCGTTINSHTTAGVVE 1726
 Db 2060 QFQVELSLRSTSEDRTPVLVVEKLINVIEMHGLYEGIRYKSGSTWKI-----K 2107
 QY 1727 KLIRGLAME-DSRNMFALFEYNGHVDKAIERTVADVLAPKELKAAISEVGLPWKRY- 1784
 Db 2108 ELRQGLDTPDASVN--LDDYNIR-----VIASVFKQW-----LRLDLPNPLMT 2147
 QY 1785 FKLY 1788
 Db 2148 FELY 2151

RESULT 15
 QRYE8
 ID QRYE8 PRELIMINARY; PRT; 1611 AA.
 AC QRYE8:
 DT 01-JUN-2002 (TREMblrel. 21, Created)
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Putative myosin heavy chain.
 GN AT2G33240.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OC NCB1_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounleley S.D.; Ketchum K.A.; Lin X.; Crosby M.L.; Brandon R.C.;
 RA Sykes S.M.; Kaul S.; Macon T.M.; Kerlavage A.R.; Adams M.D.;
 RA Somerville C.R.; Venter J.C.;
 RL Submitted (JUL-1997) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Town C.D.; Kaul S.;
 RL Submitted (FEB-2002) to the EMBL/Genbank/DBJ databases.

DR ENBL, ACC02334; AM14807.1; -
SQ SEQUENCE 1611 AA; 183741 MW; 98221A9494D60679 CRC64;

Query Match 13.9%; Score 1487.5; DB 10; Length 1611;
Best Local Similarity 32.5%; Pred. No. 8; 1e-70;
Matches 401; Conservative 194; Mismatches 378; Indels 259; Gaps 35;

QY 5 GTRVWLR-----NQHPSTVNSCAEGLVPR-TDYGQVFTYKOSTTHOK 50
DB 9 GSGVWVDPDEAMLDGEVVEANGOEI--KVCQCKTVSPSPKORDVVLTK----VWAK 62
QY 51 VTAMHPNEE-----GVDMASTLTLHGGSIMVNLFOYKKNQWTYIGSLASVNYOP 106
DB 63 VNVVHPDPEPPELGVDDMTKLAYLHPGVLNLKAYNANEIYTYGNTLLVNPFKRL 122
QY 107 AGLYEPATMEQYSRRLGELPHIIPALINCEYRCMLKMDNDGCLIKGSEAGKSTYKL 166
DB 123 PHLYGNEIMQYKGTDPGELSPHPFAVDASAYKMINEGVSOALVSGEAGKSTYK 182
QY 167 ILKFLSVISQOSELSEKTSCEVERAILESSPIMEAFGNKTVNNSSRFQVQVNI 226
DB 183 LMQYLAVMGSKA-----ESEGRSVEQOVLESNPVLEAFGNKTVNNSSRFQVQVNI 237
QY 227 CQKGNIOGRIIVDCILSSQNRVVRNGERNYHIFALLAGLEHREBEFYLPSTPENHY 286
DB 238 NHMGRISGAARITVLL-ERSRVCQSDPERNYHCFYMLCAAP-QETERYOLGKPFSTFY 295
QY 287 LNSGCEVEDTISDOESFEVITAMDVQPSKEEVESRLAGILHLNIEFI---TAG 343
DB 296 LNSGNCALDAIDSKETLTKRAMDVYGISPEQDAIFRYVAALHLGNIIEFAKSESD 355
QY 344 GAQV---SFTYALGRSALLGLDPTOLDALTORSMFLGEEILTPLVQAVDSRDSLA 400
DB 356 GAEPKDKSRPHLKVAKLFCMDEKALENSLCNRVMVTRGESITKPLDPGALSRLDAL 415
QY 401 MALYACFEVNIKINSRIGNEDFK-SIGILDPGENEVNHPFQFNINANEKLOEY 459
DB 416 KIYYSKLFDMLVTKINNSIGODSSSKYIIGVDIYGSEFKTNSFEQFCINTNEKLOOH 475
QY 460 FNKHFISLEQLEYSREGLVWEDIDWIDNGECLDLEKUL-GLLALINBSHPQATDSTL 518
DB 476 FNGHVPRMEBEETKEIDSYIEPIDNOVDLIEKKPGIITALLDEACMFPRSTHDL 535
QY 519 LEXLHSHANNHFFVYKRVAVNNFVGKRVAGEVOYDVGILEKNRDTFRDDLNLRESR 578
DB 536 AEKLYQTFGSHKFTPKLARTDFTICHYAGDVYQTELFDKNKDYVVEGHOSLMNSD 595
QY 579 FDRITYDLFEHVSSRNODTLKCGSKHRPVSQFKVDSLHSMATLSSNPPFVACIKP 638
DB 596 CSFVSSLF---PKSREBSK---SSKFSISIGQFK-QOQSLLETLNTTEPHYIRCVKP 647
QY 639 NMQKMPDQDAVVLNQLRYSGLVETVIRIKAGVAVRPFQDFYKRYKVMRLALPE-- 696
DB 648 NNVLKPEIFENNVVHLQRCGVMEAIRISCAGYPTKRPNEFLTRFIL-----APEAT 702
QY 697 ----DVAGKCTSLQLVDASNSEWOLGKTVFLRESLEQLEKREBEVSHAAMVIRAHV 752
DB 703 ERSFDEVDACKLLARVDLKG--FOIGKTVFLRAGOMELDAHRAVLSHSARIIORVY 760
QY 753 LGFLARKY-----RKY 764
DB 761 ITYLSRKVYLLLOSASTETQAFCRGHARVOFATRREASVRIQOARTYICQTAFAKL 820
QY 765 LYCVIIIOKNYRAF-----LLRRF 784
DB 821 CASAIISIGSLRAMARVEFOYRTKKAATIIQASLKPHIDDKLSFSGIRKLCRRRY 880
QY 785 LHLKKAIFYOKLQGIARVYROLLEKRE-----QEEKKQEEB-----EKKK 830
DB 881 LRTTKKAITTQCGMRYVVAHRELKLMMAKETGALQDAKTLEKEVEBELTSCLEKQW 940
QY 831 REEERERERERE-----AELRAQEEETRKQOELALOKSQKE-AELTREJ 877

DB 941 RMELDQVKTQOEVEDLRSALNDMKLOLGETQVTKSEIILKLO--SALQDMOLEFBEELAKEL 998
QY 878 EKOK-----ENKOVEEIL-RLKEKIEDL-----ORKK-----E 904
DB 999 EMTNLLAENEDLKLVLVSLQKIDESDKYEETSKLSEBRKQEVPIYDGVIIKLEAE 1058
QY 905 QOELSUTEASLOK-----LOERRDQELRLBEACCAAOFLLESNPFDE-- 948
DB 1059 NQKLALVSTLEKKIDSLDRKHDDLVDLLERKIDETEKKYEEASKLCEERLKQVVDTEK 1118
QY 949 IDECVNIERSLSGSEFSSSELAESACEKPFNFQOPYPEEEVDEGEFADDDAF----- 1003
DB 1119 YEEASRLCEERLKQVVDTEKLE-----LKTSMQRLBEKQVSD-MEAEKIIIRQOAL 1169
QY 1004 -----KDSF-----NPEHGHS 1016
DB 1170 RNSASRKMSPOKSDLVFVMTYLFQVENGHHE 1201

Search completed: July 14, 2003, 18:21:37
Job time : 122.43 secs


```

Db      1621 MARIHVKGDLSSEAMCYHVALVALEYLTRKEAVQWEPPLPHSHSACLRSRGVFQ 1680
Qy      1681 GCTAFVITPNIDEASMMEDVGMOPVFNEDVLMELLBCQADGLKARVELIADIYKL 1740
Db      1681 GCTAFVITPNIDEASMMEDVGMOPVFNEDVLMELLBCQADGLKARVELIADIYKL 1740
Db      1681 GCTAFVITPNIDEASMMEDVGMOPVFNEDVLMELLBCQADGLKARVELIADIYKL 1740
Qy      1741 IIPYHKKRDFERLAHLVDTLHRAYSKVTVMHSGRLLGTYPRVAFGQAQOYPTDSE 1800
Db      1741 IIPYHKKRDFERLAHLVDTLHRAYSKVTVMHSGRLLGTYPRVAFGQAQOYPTDSE 1800
Qy      1801 TVEGHPFEDDEKEYIYKPKLTPLSEISQRLKLYSDKFGSENMVIMIDSGKNPKDLD 1860
Db      1801 TVEGHPFEDDEKEYIYKPKLTPLSEISQRLKLYSDKFGSENMVIMIDSGKNPKDLD 1860
Qy      1861 SKYAYIQVTHVLPFPEKELOERKTEPERSHNIRRMFPMFPPTQTGKRGVBEQCKRT 1920
Db      1861 SKYAYIQVTHVLPFPEKELOERKTEPERSHNIRRMFPMFPPTQTGKRGVBEQCKRT 1920
Qy      1921 ILTALHCFPYVKRRIIPVYQHTDNLPIEVAIDENSKYAEELRQLCSAEVDMIKQLKLT 1980
Db      1921 ILTALHCFPYVKRRIIPVYQHTDNLPIEVAIDENSKYAEELRQLCSAEVDMIKQLKLT 1980
Qy      1981 QGSVSVQVAGPLAYARAFLDNTNTRYDNNKYLKEYFRQVFEACGQALAVNERLIKE 2040
Db      1981 QGSVSVQVAGPLAYARAFLDNTNTRYDNNKYLKEYFRQVFEACGQALAVNERLIKE 2040
Qy      2041 DPLEYOEMKANYREMAKELSEIMHEQICPLEEKTSLVNSLIHFAISGTPSTVWHGM 2100
Db      2041 DPLEYOEMKANYREMAKELSEIMHEQICPLEEKTSLVNSLIHFAISGTPSTVWHGM 2100
Qy      2101 TSSSSSV 2107
Db      2101 TSSSSSV 2107

```

```

PT      New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
PT      an immune response, and for treating multiple sclerosis, rheumatoid
PT      arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
PT      and sepsis
PS      Disclosure, Figure 13; 245pp; English.
XX
XX      The invention relates to an isolated polypeptide (I) comprising an amino
CC      acid sequence that has 90 % sequence identity to one of the human
CC      cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E)
CC      sequences (PS). (I) is useful for identifying a compound or agent that
CC      binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
CC      detecting a CLASP-2 polypeptide in a sample. (II) is useful for
CC      inhibiting an immune response in a subject. A pharmaceutical composition
CC      comprising a nucleic acid encoding (i), or (ii) is useful for preventing
CC      or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
CC      the autoimmune disease is caused or exacerbated by increased activity
CC      of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for
CC      detection or inhibition of CLASP-2 expression (e.g., antisense or
CC      ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
CC      polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
CC      antibodies or are used as therapeutic polypeptides. The CLASP-2
CC      polynucleotide or fragments can be used in diagnostics (e.g., as probes
CC      for CLASP-2 expression), as a lymphocyte marker and for therapeutic
CC      purposes. CLASP-2 polynucleotides can construct transgenic and knockout
CC      animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2
CC      polynucleotides can screen for CLASP-2 agonists and antagonists.
CC      CLASP-2 polypeptides or polynucleotides can treat deficiencies or
CC      disorders of the immune system, by activating or inhibiting the
CC      activation, differentiation of immune cells and can treat or detect
CC      deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
CC      or polynucleotides can increase differentiation and proliferation of
CC      haematopoietic cells, including the pluripotent stem cells to treat those
CC      disorders associated with a decrease in certain (or many) types of
CC      haematopoietic cells e.g., immunologic deficiency syndromes including
CC      blood protein disorders (e.g., agammaglobulinemia,
CC      dysgammaglobulinemia, ataxia telangiectasia, common variable
CC      immunodeficiency, DiGeorge syndrome, lymphopenia, thrombocytopenia, or
CC      haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
CC      detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
CC      Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
CC      endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
CC      inflammation. CLASP-2 can be used to treat anaphylaxis or
CC      hypersensitivity to an antigenic molecules, organ rejection or graft-
CC      versus-host disease (GVHD) and inflammation. ABG61670-ABG61708
CC      represent cadherin-like asymmetry protein (CLASP) sequences of the
CC      invention.
XX
XX      Sequence 1980 AA;
SQ
Qy      Query Match 90.8%; Score 9926.5; DB 23; Length 1980;
Db      Best Local Similarity 93.8%; Pred. No. 0;
Db      Matches 1932; Conservative 8; Mismatches 15; Indels 105; Gaps 6;
Qy      73 MLLFPYDDFOITALRROGRYICSTVPAKAEBAQSLPYTECIKTYSNOMHLVNYKEDYS 132
Db      1 MLLFPYDDFOITALRROGRYICSTVPAKAEBAQSLPYTECIKTYSNOMHLVNYKEDYS 60
Qy      133 GEFROLPNKVVKLDTLPVHYVEVDDEVDDEPDASLSGQSGITRGVLKGNNSAISV 192
Db      61 GEFROLPNKVVKLDTLPVHYVEVDDEVDDEPDASLSGQSGITRGVLKGNNSAISV 120
Qy      193 TMRSPRRRPFHLIOLGDSYKFEPLKD--LQKPKGSIPLGFLVSPNNVVRPAPFL 250
Db      121 TMRSPRRRPFHLIOLGDSYKFEPLKD--LQKPKGSIPLGFLVSPNNVVRPAPFL 179
Qy      251 KMQDKSSYLLAADSEVEHEWETITLNTKLQNFENAMQEKRGSDHEDDEQSKLGGSGG 310
Db      180 KMQDKSSYLLAADSEVEHEWETITLNTKLQNFENAMQEKRGSDHEDDEQSKLGGSGG 239
Qy      311 LDSYLPFLAKSGAREAEIKLKSGSRVLYLPDPAQKLDPSSAEPVKSFEKFGRIILV 370
Db      240 LDSYLPFLAKSGAREAEIKLKSGSRVLYLPDPAQKLDPSSAEPVKSFEKFGRIILV 299

```


XX 15-OCT-2001; 2001WO-US32202.
 PF 13-OCT-2000; 2000US-0687837.
 XX (ARBO-) ARBOR VITAE CORP.
 XX (GARW/) GARMAN J D.
 PA (CAND/) CANDIA A F.
 XX Lu PS;
 PI WPI; 2002-416861/44.
 DR N-PSDB; ABR85003.
 XX New human cadherin-like asymmetry protein(s) (CLASP)-2 for mediating
 PT an immune response, and for treating multiple sclerosis, rheumatoid
 PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
 PT and sepsis -
 XX Example 1; Figure 11; 245pp; English.
 PS The invention relates to an isolated polypeptide (I) comprising an amino
 CC acid sequence that has 90 % sequence identity to one of the human
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2a, 2b, 2c, 2e)
 CC sequences (PS). (I) is useful for identifying a compound or agent that
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for
 CC inhibiting a immune response in a subject. A pharmaceutical composition
 CC comprising a nucleic acid encoding (I), or (II) is useful for preventing
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
 CC the autoimmune disease is caused or exacerbated by increased activity
 CC of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2
 CC polynucleotides can screen for CLASP-2 agonists and antagonists.
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or
 CC disorders of the immune system, by activating or inhibiting the
 CC activation, differentiation of immune cells and can treat or detect
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
 CC or polynucleotides can increase differentiation and proliferation of
 CC haematopoietic cells, including the pluripotent stem cells to treat those
 CC disorders associated with a decrease in certain (or many) types of
 CC haematopoietic cells e.g., immunologic deficiency syndromes including
 CC blood protein disorders (e.g., agammaglobulinemia),
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable
 CC immunodeficiency, dysgeorge syndrome, lymphopenia, thrombocytopenia, or
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or
 CC hypersensitivity to an antigenic molecule, organ rejection or graft-
 CC versus-host disease (GVHD) and inflammation. ABB61670-AB61708
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the
 CC invention.
 CC Sequence 1980 AA;
 SQ

Query Match 90.7%; Score 9918.5; DB 23; Length 1980;
 Best Local Similarity 93.7%; Pred. No. 0;
 Matches 1931; Conservative 8; Mismatches 16; Indels 105; Gaps 6;

Qy 73 MLTPYDDQATILRQGRYICSTVPAKAEBAOSLPTECIKTYSWHLNLYKEDPS 132
 Db 1 MLTPYDDQATILRQGRYICSTVPAKAEBAOSLPTECIKTYSWHLNLYKEDPS 60
 Qy 133 GEFROLPNKRVKLDKLPVHYVEVDEEVDKDDAASLGSGKITKGLYKGNMNSAISV 192

Db 61 GEFROLPNKRVKLDKLPVHYVEVDEEVDKDDAASLGSGKITKGLYKGNMNSAISV 120
 Qy 193 TMRSPKRRFPHLIQGDGSYKPEFLKD--LQKEPGSIFLPGYGVSPNNKVRAPAPFL 250
 Db 121 TMRSPKRRFPHLIQGDGSYKPEFLKD--LQKEPGSIFLPGYGVSPNNKVRAPAPFL 179
 Qy 251 KMDDKSYLLAADSEVEEMWITTLNKLQNFPAAMQEKNGDSHEDBOSKLGSGSG 310
 Db 180 KMDDKSYLLAADSEVEEMWITTLNKLQNFPAAMQEKNGDSHEDBOSKLGSGSG 239
 Qy 311 LDSTYLPBLAKARAEAEITLKSBSRYKFLYLDPPDAKDLPSAEPVKSPEKFGRIIVK 370
 Db 240 LDSTYLPBLAKARAEAEITLKSBSRYKFLYLDPPDAKDLPSAEPVKSPEKFGRIIVK 299
 Qy 371 CNDLSFNLQCCVAENEBEPTTNVPEFVTLSPDIKYNRKISADPHVNLNPSVQMTAT 430
 Db 300 CNDLSFNLQCCVAENEBEPTTNVPEFVTLSPDIKYNRKISADPHVNLNPSVQMTAT 359
 Qy 431 TSPALMNGSGPETSALRGILHEAAMQYKQGISVTCPRHDIPLVARIKRYLQGSITHC 490
 Db 360 TSPALMNGSG--QSPVLKGLIHEAAMQYKQGISVTCPRHDIPLVARIKRYLQGSITHC 418
 Qy 491 AEPYKSSDSSKVAQVILNAKQACQRLGQRMPPAMAARTLFDKASGNDLNARFSAIY 550
 Db 419 AEPYKSSDSSKVAQVILNAKQACQRLGQRMPPAMAARTLFDKASGNDLNARFSAIY 478
 Qy 551 RODSNKLSNDMLKLLADPFRPEKAKLPVILGNLDTIDNVSSPFRYVNSSYPTQOF 610
 Db 479 RODSNKLSNDMLKLLADPFRPEKAKLPVILGNLDTIDNVSSPFRYVNSSYPTQOF 538
 Qy 611 ETCSKTPTTFEVEBEFPCIPKHTQPYTITNHLVYPRYLKYDSQSPAKARNIACIEF 670
 Db 539 ETCSKTPTTFEVEBEFPCIPKHTQPYTITNHLVYPRYLKYDSQSPAKARNIACIEF 598
 Qy 671 KDSDEDSQPLKCTYGRFGGVFTFSAVAALVHHQNEPDEIKIELPTQLHEKHHLL 730
 Db 599 KDSDEDSQPLKCTYGRFGGVFTFSAVAALVHHQNEPDEIKIELPTQLHEKHHLL 658
 Qy 731 TTFHVS CNSNSKSGSKRDKDVETQVGYSWLPLLDGRTVTSBOH PVASNIPSGLYOE 790
 Db 659 TTFHVS CNSNSKSGSKRDKDVETQVGYSWLPLLDGRTVTSBOH PVASNIPSGLYOE 718
 Qy 791 LGMGRHYPEIKWMDGGRPLKISTHLVSTVYTDQHLNFFOYQCKTESGQALGNELY 850
 Db 719 LGMGRHYPEIKWMDGGRPLKISTHLVSTVYTDQHLNFFOYQCKTESGQALGNELY 778
 Qy 851 KYLSLHAMEGHVMIAPLPTILNOLFRVLTATQDEAVANVTYVLIHVVAQCHEBGLSH 910
 Db 779 KYLSLHAMEGHVMIAPLPTILNOLFRVLTATQDEAVANVTYVLIHVVAQCHEBGLSH 838
 Qy 911 LRSYKVAKYKPEPVYASIKYVHELTSMITLKPSADPILTSNKLKYSWFFRVILKS 970
 Db 839 LRSYKVAKYKPEPVYASIKYVHELTSMITLKPSADPILTSNKLKYSWFFRVILKS 898
 Qy 971 MAOHLENSKYVLLENQFPASYNHAAVETVNMMLPHITQKFRDNPBASKANHSLAVFI 1030
 Db 899 MAOHLENSKYVLLENQFPASYNHAAVETVNMMLPHITQKFRDNPBASKANHSLAVFI 958
 Qy 1031 KRCEFTMDRGVFNQINNYISCPAFGDPKTLFEYKFEFLRVVCHNEHYIPLNLPMPFGK 1090
 Db 959 KRCEFTMDRGVFNQINNYISCPAFGDPKTLFEYKFEFLRVVCHNEHYIPLNLPMPFGK 1018
 Qy 1091 RIQRQODLQDYSLTDFECRNHPLVGLLRVGTALQOFREVRILAIISVKNLLIKHSFD 1150
 Db 1019 RIQRQODLQDYSLTDFECRNHPLVGLLRVGTALQOFREVRILAIISVKNLLIKHSFD 1078
 Qy 1151 DRYASRSHQARIATLYPLFGLLIENTVQRIINRVDSPPVNAQMTVKDESLAPAVNPLY 1210
 Db 1079 DRYASRSHQARIATLYPLFGLLIENTVQRIINRVDSPPVNAQMTVKDESLAPAVNPLY 1138
 Qy 1211 TPQKSTLSDNSLHKDLAGISGASPYTTSFNINSVNSADRSGLISTDSGNSLPERNS 1270

```

Db      1139 TPQKSTLDNLSLHKDLGALSGIASPYTTSTPNINSVRNDRSGSLISTSGNSLPERNS 1198
Qy      1271 EKSNSLDKHQOQSTSLGNSVVRCDKLQOSEKSLMGCFLYLKMSMDALFTYWNKASTSE 1330
Db      1199 EKSNSLDKHQOQSTSLGNSVVRCDKLQOSEKSLMGCFLYLKMSMDALFTYWNKASTSE 1258
Qy      1331 LMDPFTISEVCLHQFOYMGKRYIA-----RTGMMHARLOQL 1367
Db      1259 LMDPFTISEVCLHQFOYMGKRYIARNOEGIGFVHDRKSGQLPVSNRNGMMHARLOQL 1318
Qy      1368 SLDNSLTFFNHSYSHSDADVLHOSLLEANLATEVCLTALDTLSFTLAFKQOLLADGHP 1427
Db      1319 SLDNSLTFFNHSYSHSDADVLHOSLLEANLATEVCLTALDTLSFTLAFKQOLLADGHP 1378
Qy      1428 LMKVVDVYLCFLQKQOSETALKNVFTALRSLYKFPSTPEGRADMCALCEYLKCN 1487
Db      1379 LMKVVDVYLCFLQKQOSETALKNVFTALRSLYKFPSTPEGRADMCALCEYLKCN 1438
Qy      1488 SKLSSIRTEASQLLYFLMRNNDYTGKSFVRTHLQVILSVQLADVVGIGTRFQSL 1547
Db      1439 SKLSSIRTEASQLLYFLMRNNDYTGKSFVRTHLQVILSVQLADVVGIGTRFQSL 1498
Qy      1548 SIINNCAANDRLIKHSPSSDVVDLTKRIRTVLMAQKKEHNDPEMLVDLYSLAKSY 1607
Db      1499 SIINNCAANDRLIKHSPSSDVVDLTKRIRTVLMAQKKEHNDPEMLVDLYSLAKSY 1558
Qy      1608 ASTPELRKTWLDSEMARHYKNGDLSAANCYHVTALVLEYLRKAVOWEPPLPHSHS 1667
Db      1559 ASTPELRKTWLDSEMARHYKNGDLSAANCYHVTALVLEYLRK----- 1603
Qy      1668 ACLRSRGVFRQGTAFRVITPNIDEASMMEDVGMVHNEVDYLMELLEQCADGLWK 1727
Db      1604 -----GVFRQGTAFRVITPNIDEASMMEDVGMVHNEVDYLMELLEQCADGLWK 1655
Qy      1728 AERYELIADIYKLIITYEKRPFERLAHYDTLHAAYSKYTVWMSGRRLLGTYRVAF 1787
Db      1656 AERYELIADIYKLIITYEKRP----- 1678
Qy      1788 FGQAAQYQFTDSETVGEFFEDDGKEYIYKEPKLPLSEISORLLKLYSDKFGSENVNM 1847
Db      1679 -----FFEDDGKEYIYKEPKLPLSEISORLLKLYSDKFGSENVNM 1720
Qy      1848 IODSGKVPKDLDSKAYIQTHTVIFPDEKELQEKTEPERSHNIIRFMFEMPFTQTK 1907
Db      1721 IODSGKVPKDLDSKAYIQTHTVIFPDEKELQEKTEPERSHNIIRFMFEMPFTQTK 1780
Qy      1908 RQGVSEEOCCRRITLTAHCFPVYKRIPIWYQHTDLPPIEVALIDEMSKVAELNQLQS 1967
Db      1781 RQGVSEEOCCRRITLTAHCFPVYKRIPIWYQHTDLPPIEVALIDEMSKVAELNQLQS 1840
Qy      1968 SAEVDMIKQLKQGSVSQVNAAGPLAARAFDDTNTKRYPNKYKLKEVRFQVEAC 2027
Db      1841 SAEVDMIKQLKQGSVSQVNAAGPLAARAFDDTNTKRYPNKYKLKEVRFQVEAC 1900
Qy      2028 GQALAVNERLIKEDQLEYOEEMKANYREMAKELSEIMHBOICPLEBKTSVLPNSLHIFNA 2087
Db      1901 GQALAVNERLIKEDQLEYOEEMKANYREMAKELSEIMHBOICPLEBKTSVLPNSLHIFNA 1960
Qy      2088 ISGTPSTWVHGMTSSSSVV 2107
Db      1961 ISGTPSTWVHGMTSSSSVV 1980

```

RESULT 4

AA040315 ID AA040315 standard; Protein; 1572 AA.

AC AA040315;

DT 08-FEB-2001 (first entry)

XX Human ORFX ORF79 polypeptide sequence SEQ ID NO:158.

```

KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
KW vulnery; antipariatic; antiparkinsonian; noctropic; neuroprotective;
KW anticonvulsant; osteopathic; antidiabetic; immunosuppressive; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antidiabetic;
KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antidiabetic; antidiabetic; coagulation;
KW thrombosis; contraceptive.
OS Homo sapiens.
PN WO200058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000MO-US08621.
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
PA (CUBA-) CUBAGEN CORP.
XX Shinkets RA, Leach M,
PI WPI: 2000-602362/57.
DR N-PSDB; AAC74524.
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX Claim 11; Page 506-510; 5507pp; English.
XX AAC74446 to AAC77606 encode the proteins given in AA040315 to AA043397,
XX which represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytosolic; hepatotropic; vulnery;
XX antipariatic; antiparkinsonian; noctropic; neuroprotective;
XX osteopathic; anticonvulsant; antidiabetic; immunosuppressive;
XX immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
XX antidiabetic; hypotensive; dermatological; immunosuppressive;
XX antidiabetic; antidiabetic; antidiabetic; antidiabetic; antidiabetic;
XX antithyroid; and antianemic. The sequences can be used for determining
XX the presence of or predisposition to, or preventing or treating
XX pathological conditions associated with an ORFX-associated disorder. The
XX nucleic acids can be used to express ORFX proteins in gene therapy
XX vectors. The proteins and nucleic acids may be used to treat cancers,
XX proliferative disorders, neurodegenerative disorders, osteoarthritis,
XX graft vs host disease, cardiovascular disease, diabetes mellitus,
XX hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
XX erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
XX bacterial or fungal infection, malaria, autoimmune disorders, asthma,
XX allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
XX nocturnal haemoglobinuria, antidiabetic disease; to enhance
XX coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 1572 AA;

```

Query Match 74.7%; Score 8173; DB 21; Length 1572;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 1571; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 536 ASGNLDKARPSATYRONSNTLSNDMLKLADPKPKPMKLYPIYGLNLTDTINVSDD 595
1 ASGNLDKARPSATYRONSNTLSNDMLKLADPKPKPMKLYPIYGLNLTDTINVSDD 60

Qy 566 FNNVNSSYIPTKQETCSKTPTIEVEEFVPCIPKHTOPTITITNHLVYVRYKYLYKDSQ 655
 Db 61 FNNVNSSYIPTKQETCSKTPTIEVEEFVPCIPKHTOPTITITNHLVYVRYKYLYKDSQ 120
 Qy 656 KSPAKARNALICIEPRDSEEDSOPKCIYGRPGGVFTRSAFAVLAHHNQPEFDEK 715
 Db 121 KSPAKARNALICIEPRDSEEDSOPKCIYGRPGGVFTRSAFAVLAHHNQPEFDEK 180
 Qy 716 IELPTQHEKHLHLPTFFHVSQNSKSGTKKRDVETOVGYSMPLDKGVSMTSEOH 775
 Db 181 IELPTQHEKHLHLPTFFHVSQNSKSGTKKRDVETOVGYSMPLDKGVSMTSEOH 240
 Qy 776 PVSANI,PSGYLYQELGMRHAYGPEIKWDGSKPLKISTHLSVTVYTOQDHLNFPQYC 835
 Db 241 PVSANI,PSGYLYQELGMRHAYGPEIKWDGSKPLKISTHLSVTVYTOQDHLNFPQYC 300
 Qy 836 QRTBSGAOLGNELYKYLKSLHAMEGHVIAFLPTLLNQLFRVLTTRATOEBAVAVNTRYI 895
 Db 301 QRTBSGAOLGNELYKYLKSLHAMEGHVIAFLPTLLNQLFRVLTTRATOEBAVAVNTRYI 360
 Qy 896 IHVVAQCHBEGLESNLSRYKYKAYKAPYVASEYKTVHEILTKSMTTILKPSADFLTSNK 955
 Db 361 IHVVAQCHBEGLESNLSRYKYKAYKAPYVASEYKTVHEILTKSMTTILKPSADFLTSNK 420
 Qy 956 LKYSWFFEDVLIKSMAGHLIENSKYKLRNORFPAHYHAYETVVMMLMPHITOKFRDN 1015
 Db 421 LKYSWFFEDVLIKSMAGHLIENSKYKLRNORFPAHYHAYETVVMMLMPHITOKFRDN 480
 Qy 1016 PRASKANSLSAVFIKRCCTFMDRGFVFOQINNYISCFAPGDPKTLFEKXFEFLRVCMH 1075
 Db 481 PRASKANSLSAVFIKRCCTFMDRGFVFOQINNYISCFAPGDPKTLFEKXFEFLRVCMH 540
 Qy 1076 EHYIFLNLMPMPGKRIORVODLQDLSLTDFECRHHFVGLILREVGTALEPFREVRLL 1135
 Db 541 EHYIFLNLMPMPGKRIORVODLQDLSLTDFECRHHFVGLILREVGTALEPFREVRLL 600
 Qy 1136 AISVLKNLILKHSFDRVYASRSHQARIATLYPLFGLLIENYQRIINVRDVSPPVNAGMT 1195
 Db 601 AISVLKNLILKHSFDRVYASRSHQARIATLYPLFGLLIENYQRIINVRDVSPPVNAGMT 660
 Qy 1196 VVDESLALPANNPLTTPQKSTLSDNSLHDLGALSGIASPTTTSPPININSRNADSRGS 1255
 Db 661 VVDESLALPANNPLTTPQKSTLSDNSLHDLGALSGIASPTTTSPPININSRNADSRGS 720
 Qy 1256 LISTSGNSLPERNSEKNSLKHQOOSTLGSVVRCDLDOSEIKSLMCFYLYTKSMS 1315
 Db 721 LISTSGNSLPERNSEKNSLKHQOOSTLGSVVRCDLDOSEIKSLMCFYLYTKSMS 780
 Qy 1316 DDALFTYNNKASTSEIMDPFTISEVCLHOFQYMGKRYIARTGMMHARLQOLGSLDNLSTF 1375
 Db 781 DDALFTYNNKASTSEIMDPFTISEVCLHOFQYMGKRYIARTGMMHARLQOLGSLDNLSTF 840
 Qy 1376 NNSYGHSDADVTHOSILEANITTEVCLNLDLTLSTFLAFAKQOLADHGNPLMKKVPDV 1435
 Db 841 NNSYGHSDADVTHOSILEANITTEVCLNLDLTLSTFLAFAKQOLADHGNPLMKKVPDV 900
 Qy 1436 YLCPLOKHOSETALKVFTALSLIYKPESTFEGRADCALCYEILKCNKSLSTRT 1495
 Db 901 YLCPLOKHOSETALKVFTALSLIYKPESTFEGRADCALCYEILKCNKSLSTRT 960
 Qy 1496 BASQILLYFLMRNPNFYTGKKSFEVTHLQYIISVSQILACVVGIGTRFOQSLIINNCA 1555
 Db 961 BASQILLYFLMRNPNFYTGKKSFEVTHLQYIISVSQILACVVGIGTRFOQSLIINNCA 1020
 Qy 1556 SBRLLKHTSFSSDVXDLTKRITVLMATQMKHEHNDPEMLVDLOYSLAKSYASTPELAK 1615
 Db 1021 SBRLLKHTSFSSDVXDLTKRITVLMATQMKHEHNDPEMLVDLOYSLAKSYASTPELAK 1080
 Qy 1616 TWLDSMARLHVNGGDLSEAMCYVHTALVAEYLTKKEVQVMBPPLPHSHSACLRSSRG 1675
 Db 1081 TWLDSMARLHVNGGDLSEAMCYVHTALVAEYLTKKEVQVMBPPLPHSHSACLRSSRG 1140
 Qy 1676 GVFRQCTAFRVITPNIIDEASMMEDVGMQDVHFNEVDIMELLBOCADGLMRAERYELIA 1735

Db 1141 GVFRQCTAFRVITPNIIDEASMMEDVGMQDVHFNEVDIMELLBOCADGLMRAERYELIA 1200
 Qy 1736 DIYKLIITIEYKRRDPEERLALYDTLHRAVSKVTEVMSGRLLIGTVFVAFFGQAQYQ 1795
 Db 1201 DIYKLIITIEYKRRDPEERLALYDTLHRAVSKVTEVMSGRLLIGTVFVAFFGQAQYQ 1260
 Qy 1796 FTDSEYVEGFPEDDGKEYIYKEPKLTPLSIESQRLKLYSDKFGSENVKMIQDSKYN 1855
 Db 1261 FTDSEYVEGFPEDDGKEYIYKEPKLTPLSIESQRLKLYSDKFGSENVKMIQDSKYN 1320
 Qy 1856 PKDLSKAYIYQVTHVIFPPEKEILOEKTEPERSHNIIRFMPFMPFTOTGKROGVBEQ 1915
 Db 1321 PKDLSKAYIYQVTHVIFPPEKEILOEKTEPERSHNIIRFMPFMPFTOTGKROGVBEQ 1380
 Qy 1916 CKRRTILAHICFPYVKRIPVMYOHNDLNDPIEVAIDEMSKVAELROLCSAABVDIK 1975
 Db 1381 CKRRTILAHICFPYVKRIPVMYOHNDLNDPIEVAIDEMSKVAELROLCSAABVDIK 1440
 Qy 1976 LQKLQGSVSVQVNAGPLAVARAFLDDTNTKRYPNKVKLLKEVFRQFVACGQALAVNE 2035
 Db 1441 LQKLQGSVSVQVNAGPLAVARAFLDDTNTKRYPNKVKLLKEVFRQFVACGQALAVNE 1500
 Qy 2036 RLKEDQLEYOEMKANYRENAKELSEIMHQICPLEBKTSVLPNSLHIFNAISGTPST 2095
 Db 1501 RLKEDQLEYOEMKANYRENAKELSEIMHQICPLEBKTSVLPNSLHIFNAISGTPST 1560
 Qy 2096 MVHGMTSSSSVY 2107
 Db 1561 MVHGMTSSSSVY 1572

RESULT 5
 ID ABG61676 standard; Protein; 1534 AA.
 XX
 AC ABG61676;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Cadherin-like asymmetry protein (CLASP) isoform #5.
 XX
 XX Human; autoimmune disease; haematopoietic disorder; Digesorge syndrome;
 KW blood protein disorder; agammaglobulinaemia; dysgammaglobulinaemia;
 KW ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
 KW thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
 KW haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;
 KW endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
 KW autoimmune pulmonary inflammation; organ rejection; inflammation;
 KW CLASP.
 XX
 OS Homo sapiens.
 XX
 PN WO200231117-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 15-OCT-2001; 2001WO-US32202.
 XX
 PR 13-OCT-2000; 2000US-0687837.
 XX
 PA (ARBO-) ARBOR VITA CORP.
 PA (GARM-) GARMAN J D.
 PA (CAND/) CANDIA A F.
 XX
 PI Lu PS;
 XX
 DR WPI: 2002-416861/44.
 XX
 PT N-PSDB: ABK84970.
 PT New human cadherin-like asymmetry protein(e) (CLASP)-2 for modulating
 an immune response, and for treating multiple sclerosis, rheumatoid
 arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,

PT and sepsis -
 XX
 PS Disclosure; Figure 3B; 245pp; English.
 CC
 CC The invention relates to an isolated polypeptide (I) comprising an amino
 XX acid sequence that has 90 % sequence identity to one of the human
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2a, 2b, 2C, 2E)
 CC sequences (PS). (I) is useful for identifying a compound or agent that
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for
 CC inhibiting a immune response in a subject. A pharmaceutical composition
 CC comprising a nucleic acid encoding (I) or (II) is useful for preventing
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
 CC the autoimmune disease is caused or exacerbated by increased activity
 CC of Th1 cells. CLASP-2 polynucleotides are useful as probes or primers for
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists.
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or
 CC disorders of the immune system, by activating or inhibiting the
 CC activation, differentiation of immune cells and can treat or detect
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
 CC or polynucleotides can increase differentiation and proliferation of
 CC haematopoietic cells, including the pluripotent stem cells to treat those
 CC disorders associated with a decrease in certain (or many) types of
 CC haematopoietic cells e.g., immunologic deficiency syndromes including
 CC blood protein disorders (e.g., agammaglobulinemia,
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable
 CC immunodeficiency, DiGeorge syndrome, lymphopenia, thrombocytopenia, or
 CC detect autoimmune diseases, e.g., Addison's disease, hemolytic anaemia,
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or
 CC hypersensitivity to an antigenic molecules, organ rejection or graft-
 CC versus-host disease (GVHD) and inflammation. AbG61670-ABG61708
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the
 CC invention.
 CC
 XX
 XX
 SQ Sequence 1534 AA;
 Query Match 72.9%; Score 7976; DB 23; Length 1534;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 301 QKTESGAQALGNELVLYKSLHAMEGHVIAFLPTILNQLFRVLTRATOEEVAVNTRVI 360
 QY IHVVAQCHBEGLSHLSRYVYKAYAPYVASEYTVHEELTKSMTTILKPSADLTSENK 955
 DB 896
 QY IHVVAQCHBEGLSHLSRYVYKAYAPYVASEYTVHEELTKSMTTILKPSADLTSENK 420
 DB 361
 QY LKYSWFFFDVLIKSMAOHLIENSKVLLRLRQPPASVHHAVETVNMIMPHITOKFRDN 1015
 DB 956
 QY LKYSWFFFDVLIKSMAOHLIENSKVLLRLRQPPASVHHAVETVNMIMPHITOKFRDN 480
 DB 421
 QY LKYSWFFFDVLIKSMAOHLIENSKVLLRLRQPPASVHHAVETVNMIMPHITOKFRDN 480
 DB 481
 QY PEASNAHNSIAVFIKRCFTFMDRGFVFKQINNYISCAPGPKTLFEYKEFEFLVNCN 540
 DB 1016
 QY PEASNAHNSIAVFIKRCFTFMDRGFVFKQINNYISCAPGPKTLFEYKEFEFLVNCN 1075
 DB 481
 QY PEASNAHNSIAVFIKRCFTFMDRGFVFKQINNYISCAPGPKTLFEYKEFEFLVNCN 540
 DB 1076
 QY EHYIPLNLPMPGKRIRYODLODYSLTDFECNHNFLVGLLREVGTLAOEPEFVRLI 1135
 DB 541
 QY EHYIPLNLPMPGKRIRYODLODYSLTDFECNHNFLVGLLREVGTLAOEPEFVRLI 600
 DB 1136
 QY AIVLKNLILIKHSFDDRYSRSHQARIATVLPFGLLIENVQINVDVSPFVNAQMT 1195
 DB 601
 QY AIVLKNLILIKHSFDDRYSRSHQARIATVLPFGLLIENVQINVDVSPFVNAQMT 660
 DB 1136
 QY VDESLALPAVNPVLTPOKSTLDSNLKDLGALISGASPYTSTPINSVRNADSRGS 1255
 DB 661
 QY VDESLALPAVNPVLTPOKSTLDSNLKDLGALISGASPYTSTPINSVRNADSRGS 720
 DB 1256
 QY LISTSGNSLPERNSEKNSLDKHOQSTLGSVYRCDKLOOSEIKSLMCFYLTKSMS 1315
 DB 721
 QY LISTSGNSLPERNSEKNSLDKHOQSTLGSVYRCDKLOOSEIKSLMCFYLTKSMS 780
 DB 1316
 QY DDALEFTYNNKASTSELMDFITISEVCLHQFYMGKRYIARTGMHARLQOLGSLDNLSTF 1375
 DB 781
 QY DDALEFTYNNKASTSELMDFITISEVCLHQFYMGKRYIARTGMHARLQOLGSLDNLSTF 840
 DB 1376
 QY NNSYGHSDADVLHOSLEENANITEVCLTALDTLSPTAFKQOLLADGHNPLMKVVDV 1435
 DB 841
 QY NNSYGHSDADVLHOSLEENANITEVCLTALDTLSPTAFKQOLLADGHNPLMKVVDV 900
 DB 1436
 QY YICFLQKQOSEFALKNVFTALRSLLYKPPSTFEBRADMCALCEYILKCCNSKLSIRT 1495
 DB 901
 QY YICFLQKQOSEFALKNVFTALRSLLYKPPSTFEBRADMCALCEYILKCCNSKLSIRT 960
 DB 1496
 QY EASQLLYFLMRNPFDTYKSKSFVRTHLQVVISVSQLADVVGIGTRFQOQSLIINNCAN 1555
 DB 961
 QY EASQLLYFLMRNPFDTYKSKSFVRTHLQVVISVSQLADVVGIGTRFQOQSLIINNCAN 1020
 DB 1556
 QY SDRLIKHTSFSSDVVDLTKRIRTVLMATQKHEHNDPEMLVDLOYSLAKSYASTPELRK 1615
 DB 1021
 QY SDRLIKHTSFSSDVVDLTKRIRTVLMATQKHEHNDPEMLVDLOYSLAKSYASTPELRK 1080
 DB 1616
 QY TWLDSMARIHVKNGLSEAMACVYHTALVAEYLTRKSAVQWEPPLPHSHSACLRSRG 1675
 DB 1081
 QY TWLDSMARIHVKNGLSEAMACVYHTALVAEYLTRKSAVQWEPPLPHSHSACLRSRG 1140
 DB 1676
 QY GVFRQCTAFRVITBNIDEASMMEDVGMQDVHFNEDVJMBLEQCADGLMAERYELIA 1735
 DB 1141
 QY GVFRQCTAFRVITBNIDEASMMEDVGMQDVHFNEDVJMBLEQCADGLMAERYELIA 1200
 DB 1736
 QY DIYKILIIYERKRPFERLAHLYTLRAYSKYTVHWSGRRLLCTYRVVAFPGQAAQY 1795
 DB 1201
 QY DIYKILIIYERKRPFERLAHLYTLRAYSKYTVHWSGRRLLCTYRVVAFPGQAAQY 1260
 DB 1796
 QY FTDSEYVGEFPEDDGKEYIYKEPKLPLLSISORLLKLVSDKGSENVKMIQDSGKN 1855
 DB 1261
 QY FTDSEYVGEFPEDDGKEYIYKEPKLPLLSISORLLKLVSDKGSENVKMIQDSGKN 1320
 DB 1856
 QY PKDLSKAYAIQVTHVIFPDEKELQERKTEBERSHNIIRFMFEMPTQGRQGVBEQ 1915
 DB 1321
 QY PKDLSKAYAIQVTHVIFPDEKELQERKTEBERSHNIIRFMFEMPTQGRQGVBEQ 1380
 DB 1916
 QY CKRRTILTAIHCFPVYKRIIPWVOHTDNLPIEVAIDEMSKVAELQCSSAAVDNIK 1975
 DB 1381
 QY CKRRTILTAIHCFPVYKRIIPWVOHTDNLPIEVAIDEMSKVAELQCSSAAVDNIK 1440

QY 1976 10LKLQGSVSVQVNNAGPLAVYAPAFDDITNTKRYPDNKKVLLKLVPRQFYACGQALAVNE 2035
 DB 1441 10LKLQGSVSVQVNNAGPLAVYAPAFDDITNTKRYPDNKKVLLKLVPRQFYACGQALAVNE 1500
 QY 2036 RLKKEPOLLEYOEEMKANYEMAKELSEIMHEOI 2068
 DB 1501 RLKKEPOLLEYOEEMKANYEMAKELSEIMHEOI 1533

RESULT 6
 ABG61680
 ID ABG61680 standard, Protein, 1534 AA.
 AC ABG61680;
 DT 13-AUG-2002 (first entry)
 XX
 DE Cadherin-like asymmetry protein (CLASP) isoform #9.
 XX
 KW Human; autoimmune disease; haematopoietic disorder; Digorge syndrome;
 KW blood protein disorder; agammaglobulinaemia; dysgammaglobulinaemia;
 KW ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
 KW chromocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
 KW haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;
 KW endometritis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
 KW autoimmune pulmonary inflammation; organ rejection; inflammation;
 KW CLASP.
 XX
 OS Homo sapiens.
 XX
 PN WO20023117-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 15-OCT-2001; 2001WO-US32202.
 XX
 PR 13-OCT-2000; 2000US-0687837.
 XX
 PA (ARBRO-) ARBOR VITAE CORP.
 PA (GARM-) GARMAN J D.
 PA (CAND/) CANDIA A P.
 XX
 PI Lu PI;
 XX
 DR WPI; 2002-416861/44.
 XX
 PT New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
 PT an immune response, and for treating multiple sclerosis, rheumatoid
 PT arthritis, endometritis, lupus, autoimmune thyroiditis, septic shock,
 PT and sepsis
 XX
 PS Disclosure; Figure 5A; 245pp; English.
 XX
 CC The invention relates to an isolated polypeptide (I) comprising an amino
 CC acid sequence that has 90 % sequence identity to one of the human
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2D)
 CC sequences (PS). (I) is useful for identifying a compound or agent that
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for
 CC inhibiting a immune response in a subject. A pharmaceutical composition
 CC comprising a nucleic acid encoding (I), or (II) is useful for preventing
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
 CC the autoimmune disease is caused or exacerbated by increased activity
 CC of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2

CC polynucleotides can screen for CLASP-2 agonists and antagonists.
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or
 CC disorders of the immune system, by activating or inhibiting the
 CC activation, differentiation of immune cells and can treat or detect
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
 CC or polynucleotides can increase differentiation and proliferation of
 CC haematopoietic cells, including the pluripotent stem cells to treat those
 CC disorders associated with a decrease in certain (or many) types of
 CC haematopoietic cells e.g., immunologic deficiency syndromes including
 CC blood protein disorders (e.g., agammaglobulinaemia,
 CC dysgammaglobulinaemia, ataxia telangiectasia, common variable
 CC immunodeficiency, Digorge syndrome, lymphopenia, chromocytopenia, or
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
 CC endometritis, autoimmune thyroiditis, and autoimmune pulmonary
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or
 CC hypersensitivity to an antigenic molecules, organ rejection or graft-
 CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61708
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the
 CC invention.

Sequence 1534 AA;
 SQ
 Query Match 72.9%; Score 7976; DB 23; Length 1534;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 536 ASGNLDKARPSAIIYRQSNKLSNDMLKLADFKPKPKMAKLPYLGNLDTITNVSSD 595
 DB 1 ASGNLDKARPSAIIYRQSNKLSNDMLKLADFKPKPKMAKLPYLGNLDTITNVSSD 60
 QY 596 PPNVNSSYIPIKQETCSKTIPTPEVEEFPVPCIPKHQPTIYNHLYVVPKYLYQSO 655
 DB 61 PPNVNSSYIPIKQETCSKTIPTPEVEEFPVPCIPKHQPTIYNHLYVVPKYLYQSO 120
 QY 656 KSFARARNAICIEFKDSDEDSQPLKCIYGRPGSPVTRSAFAVLAHHQNPBYDRIK 715
 DB 121 KSFARARNAICIEFKDSDEDSQPLKCIYGRPGSPVTRSAFAVLAHHQNPBYDRIK 180
 QY 716 IELPQLHEKHHLLTFPHVSGDNSSKSGTKKRDVETQVGYSMPLKDKGRVVTSECHI 775
 DB 181 IELPQLHEKHHLLTFPHVSGDNSSKSGTKKRDVETQVGYSMPLKDKGRVVTSECHI 240
 QY 776 PVSANLPBGYGYOELGKGRHYGPEIKWVDGKPLKTI STLHVSVTYQDQHLNFPQYC 835
 DB 241 PVSANLPBGYGYOELGKGRHYGPEIKWVDGKPLKTI STLHVSVTYQDQHLNFPQYC 300
 QY 836 QKTBSGAQALGNELVKYLSLHAMEGHVMIAPLPTILNLFRVLTFRATQEBVAVNTRVI 895
 DB 301 QKTBSGAQALGNELVKYLSLHAMEGHVMIAPLPTILNLFRVLTFRATQEBVAVNTRVI 360
 QY 896 IHVVAQCHEBGLSHLSRYKYAYKAEPYVASEYKTHBEILTKSMTTILKPSADFLTSNK 955
 DB 361 IHVVAQCHEBGLSHLSRYKYAYKAEPYVASEYKTHBEILTKSMTTILKPSADFLTSNK 420
 QY 956 LKTSWPFEDVLISMAOHLIENSKVKLLRNORPASVHHAVETVNNMLMPTITQKPDN 1015
 DB 421 LKTSWPFEDVLISMAOHLIENSKVKLLRNORPASVHHAVETVNNMLMPTITQKPDN 480
 QY 1016 PEASKNANSHLAVFIKRCFTMDRGFVKQJNNYISCPAPDPKTLPEYKEEFLVNCNH 1075
 DB 481 PEASKNANSHLAVFIKRCFTMDRGFVKQJNNYISCPAPDPKTLPEYKEEFLVNCNH 540
 QY 1076 EHYIPLNIMPFGKRIORYODLDVSLTDFECRNHFLVGLLREVGTAQOEPRVRLI 1135
 DB 541 EHYIPLNIMPFGKRIORYODLDVSLTDFECRNHFLVGLLREVGTAQOEPRVRLI 600
 QY 1136 AIVSVKULLIHSPPDRASSHOARIATLPLFGILLIENVORINVDVSPFPNACMT 1195
 DB 601 AIVSVKULLIHSPPDRASSHOARIATLPLFGILLIENVORINVDVSPFPNACMT 660
 QY 1196 VKDESLALPAVNPVTPQKGSITLNSLHKDLGASIGIASPYTTSTPNINSVRNADSRGS 1255

Db 661 VDESEALPAVPLVPOKSGTLDNSHDLGALISGASPYTTSPTNINSVNMDSRS 720
 1256 LISTDSGNSLPERNSKSNLDRHQOSTLGNVANCXLDJOSEISLMLCFYLKXNS 1315
 Db 721 LISTDSGNSLPERNSKSNLDRHQOSTLGNVANCXLDJOSEISLMLCFYLKXNS 780
 Qy 1316 DDALFTYNNKASTSELMDFETISEVCLHQFOYMKRYIARTGMHARLOOLGSLDNLTF 1375
 Db 781 DDALFTYNNKASTSELMDFETISEVCLHQFOYMKRYIARTGMHARLOOLGSLDNLTF 840
 Qy 1376 NNSYGSADADVHLQSLLEANIATEVCLTALDITSLFTLAPKQOLADHGNPLMKVFPV 1435
 Db 841 NNSYGSADADVHLQSLLEANIATEVCLTALDITSLFTLAPKQOLADHGNPLMKVFPV 900
 Qy 1436 YICFLQKQSEPLAKVFTALRLSLIKPSTFEGRADCAALCYEILKCNKSLSTRT 1495
 Db 901 YICFLQKQSEPLAKVFTALRLSLIKPSTFEGRADCAALCYEILKCNKSLSTRT 960
 Qy 1496 EASQLLYFLMRNNDYTGKKSFVRTHLOVYISVSQIADVVGIGTRFOOSLSIINNCAN 1555
 Db 961 EASQLLYFLMRNNDYTGKKSFVRTHLOVYISVSQIADVVGIGTRFOOSLSIINNCAN 1020
 Qy 1556 SDRLLKHTSPSSVVKLTIRIVLMAATQMKHEHNDPEMLVDLOSLAKSYASTPELRK 1615
 Db 1021 SDRLLKHTSPSSVVKLTIRIVLMAATQMKHEHNDPEMLVDLOSLAKSYASTPELRK 1080
 Qy 1616 TMLDSMARLHVKNQDSEAMCYVHTALVAEYLTRKEAVQMPPLPHSHSCLRRSG 1675
 Db 1081 TMLDSMARLHVKNQDSEAMCYVHTALVAEYLTRKEAVQMPPLPHSHSCLRRSG 1140
 Qy 1676 GVFRQCTAFRVITPNIIDEBSMMEVDVGMQDVHFNEDVIMLELQCADGLMKRERYELIA 1735
 Db 1141 GVFRQCTAFRVITPNIIDEBSMMEVDVGMQDVHFNEDVIMLELQCADGLMKRERYELIA 1200
 Qy 1736 DIYKLIPIYEKRDDERLAHLVDTLHRAYSKTEVMSGRRLGTYFRVAFPGQAQOY 1795
 Db 1201 DIYKLIPIYEKRDDERLAHLVDTLHRAYSKTEVMSGRRLGTYFRVAFPGQAQOY 1260
 Qy 1796 FPDSEVDGFFDEDEGKEYIYKPLTLPSEISQRLKLYSDKFSSENVKMIQDSGKN 1855
 Db 1261 FPDSEVDGFFDEDEGKEYIYKPLTLPSEISQRLKLYSDKFSSENVKMIQDSGKN 1320
 Qy 1856 PKDLDSKVAIYQVTHVIPFDEKELQERKTEPERSHNIRPFMEPFTQGRKQGVERO 1915
 Db 1321 PKDLDSKVAIYQVTHVIPFDEKELQERKTEPERSHNIRPFMEPFTQGRKQGVERO 1380
 Qy 1916 CKRRTILTALHCPYKRIIPWYQHNTDNLPIEVAIDEMSKVAALRQLCSSAEVDMTK 1975
 Db 1381 CKRRTILTALHCPYKRIIPWYQHNTDNLPIEVAIDEMSKVAALRQLCSSAEVDMTK 1440
 Qy 1976 LQKLOGSVSVQVNAFLAYARAFLDPTTKRYPDKVKVLLKEVFQFVACQALAVNE 2035
 Db 1441 LQKLOGSVSVQVNAFLAYARAFLDPTTKRYPDKVKVLLKEVFQFVACQALAVNE 1500
 Qy 2036 RLKEDQLEQEBMKANYREMAKELSEIMHBOJ 2068
 Db 1501 RLKEDQLEQEBMKANYREMAKELSEIMHBOJ 1533

KW chromocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
 KW haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;
 KW endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
 KW autoimmune pulmonary inflammation; organ rejection; inflammation;
 KW CLASP.
 OS Homo sapiens.
 XX WO200231117-A2.
 XX 18-APR-2002.
 PD 15-OCT-2001; 2001WO-US32202.
 PF 13-OCT-2000; 2000US-0687837.
 PR (ARBO-) ARBOR VITA CORP.
 PA (GARM/) GARMAN J D.
 PA (CAND/) CANDIA A F.
 PI Lu PS;
 XX WPI; 2002-416861/44.
 DR New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
 XX an immune response, and for treating multiple sclerosis, rheumatoid
 PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
 PT and sepsis -
 PS Example 4; Figure 6B; 245pp; English.
 XX The invention relates to an isolated polypeptide (I) comprising an amino
 CC acid sequence that has 90 % sequence identity to one of the human
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E)
 CC sequences (PS). (I) is useful for identifying a compound or agent that
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for
 CC inhibiting a immune response in a subject. A pharmaceutical composition
 CC comprising a nucleic acid encoding (I), or (II) is useful for preventing
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
 CC the autoimmune disease is caused or exacerbated by increased activity
 CC of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout
 CC animals, e.g., for screening for CLASP-2 agonists and antagonists. CLASP-2
 CC polynucleotides can screen for CLASP-2 agonists and antagonists.
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or
 CC disorders of the immune system, by activating or inhibiting the
 CC activation, differentiation of immune cells and can treat or detect
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
 CC or polynucleotides can increase differentiation and proliferation of
 CC haematopoietic cells, including the pluripotent stem cells to treat those
 CC disorders associated with a decrease in certain (or many) types of
 CC haematopoietic cells e.g., immunologic deficiency syndromes including
 CC blood protein disorders (e.g., agammaglobulinemia,
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable
 CC immunodeficiency, DiGeorge syndrome, lymphopenia, chromocytopenia, or
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or
 CC hypersensitivity to an antigenic molecules, organ rejection or graft-
 CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61708
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the
 CC invention.
 XX Sequence 1534 AA;
 SQ

Query Match 72.9%; Score 7976; DB 23; Length 1534;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

QY 536 ASGNLDKARFSAIIRKODSNKLSNDMLKLLADFKREBKAKLPVTLGNLDTIDNVSSD 595
DB 1 ASGNLDKARFSAIIRKODSNKLSNDMLKLLADFKREBKAKLPVTLGNLDTIDNVSSD 60
QY 596 PPNVNVSSYIPTRKOPETCKPTTFEVEBFVPCIPGHTQYITTYTNHLVYVKYKQYSD 655
DB 61 PPNVNVSSYIPTRKOPETCKPTTFEVEBFVPCIPGHTQYITTYTNHLVYVKYKQYSD 120
QY 656 KSPAKARNIAICIEFKDSEDEDSQPLKCIYGRPGCVFTRSAFAVALHNNQDEFEYDEIK 715
DB 121 KSPAKARNIAICIEFKDSEDEDSQPLKCIYGRPGCVFTRSAFAVALHNNQDEFEYDEIK 180
QY 716 IELPTQLHEKHHLLTFPHVSCDNSKSGSTKKRDVETQYGYSMPLKDGKRVVTSQHI 775
DB 181 IELPTQLHEKHHLLTFPHVSCDNSKSGSTKKRDVETQYGYSMPLKDGKRVVTSQHI 240
QY 776 PVSANLPSGGLGYOELGMRHYPGLIKWTDGKPKLKISTHUVSTVYTDQHLNFFQYC 835
DB 241 PVSANLPSGGLGYOELGMRHYPGLIKWTDGKPKLKISTHUVSTVYTDQHLNFFQYC 300
QY 836 QKTESGAQALGNELVYKLSLHAMEGHWIAFLPTLLNQLFRVLTGATOEVAVAVNTVRI 895
DB 301 QKTESGAQALGNELVYKLSLHAMEGHWIAFLPTLLNQLFRVLTGATOEVAVAVNTVRI 360
QY 896 IHVVAQCHEGLESHRSYKCAKYKAPYVASEYKVTHEELTKSMWTTIKPSADFLTSNK 955
DB 361 IHVVAQCHEGLESHRSYKCAKYKAPYVASEYKVTHEELTKSMWTTIKPSADFLTSNK 420
QY 956 LKXWMEFFPDVLIKMAOHLIENSKYKLLRNORFPAHYHAEVTVNMLMPHITQKERN 1015
DB 421 LKXWMEFFPDVLIKMAOHLIENSKYKLLRNORFPAHYHAEVTVNMLMPHITQKERN 480
QY 1016 PEASIGNANSLAVFIKRCCTFMDRGFVFKQINNYSICPAFGPKTLFEYKFEPLRVACNH 1075
DB 481 PEASIGNANSLAVFIKRCCTFMDRGFVFKQINNYSICPAFGPKTLFEYKFEPLRVACNH 540
QY 1076 EHYIPLNLPMPFGKRIQRVQDLQDLSLTDEFGRNHPVGLLLREVGTALEFREVRLLI 1135
DB 541 EHYIPLNLPMPFGKRIQRVQDLQDLSLTDEFGRNHPVGLLLREVGTALEFREVRLLI 600
QY 1136 AHSVKNLIIKHSFDRVARSRSHQARIATLYPLFGLLENVQRLNVRVSPPPVNAAGMT 1195
DB 601 AHSVKNLIIKHSFDRVARSRSHQARIATLYPLFGLLENVQRLNVRVSPPPVNAAGMT 660
QY 1196 VDESLALPAAVNPVLPQKGSTLDNSLHDLGAIASPIYTTSTPININSVRNADSRGS 1255
DB 661 VDESLALPAAVNPVLPQKGSTLDNSLHDLGAIASPIYTTSTPININSVRNADSRGS 720
QY 1256 LISTSQNSLPERNSEKNSLDKQOQSTLGNSVACDLKDSEIKSLMCEPLYILKNS 1315
DB 721 LISTSQNSLPERNSEKNSLDKQOQSTLGNSVACDLKDSEIKSLMCEPLYILKNS 780
QY 1316 DDALFTYNNKASTSELMDFITISEVCLHQPYWNGKRYIARTGMARLQOLGSLDNLSTLF 1375
DB 781 DDALFTYNNKASTSELMDFITISEVCLHQPYWNGKRYIARTGMARLQOLGSLDNLSTLF 840
QY 1376 NNSYGHSDADVLAHQSLLEANIATEVCLTALDTLSLFTLAFKQQLADHGNPLMKVCFDV 1435
DB 841 NNSYGHSDADVLAHQSLLEANIATEVCLTALDTLSLFTLAFKQQLADHGNPLMKVCFDV 900
QY 1436 YLCFLOKHOSETALKNVFTALNSLIYKPSYFEGHADMCAALCYEILKCCNSKLSIRT 1495
DB 901 YLCFLOKHOSETALKNVFTALNSLIYKPSYFEGHADMCAALCYEILKCCNSKLSIRT 960
QY 1496 EASQILYFLMRNPDVTGKSPFRTHLQVYIISVQLIADVNGIGTRFOQSLSTINNCAN 1555
DB 961 EASQILYFLMRNPDVTGKSPFRTHLQVYIISVQLIADVNGIGTRFOQSLSTINNCAN 1020

```

```

QY 1556 SDRLIKHTSPSSDVADLTIRKIRITVIMATAQMKHEHNDPEMLVDLQYSLAKSYASTPELRK 1615
DB 1021 SDRLIKHTSPSSDVADLTIRKIRITVIMATAQMKHEHNDPEMLVDLQYSLAKSYASTPELRK 1080
QY 1616 TWLDSMAIIVHNGNGLSEAAKCYAVNTALVAEYLTIRKKAQVQMBPPLPHSHSACLRRSRG 1675
DB 1081 TWLDSMAIIVHNGNGLSEAAKCYAVNTALVAEYLTIRKKAQVQMBPPLPHSHSACLRRSRG 1140
QY 1676 GVFRQGCIAFRVITNNIDEEASNMEDVGMQDVHFNEVDYLMELLBOCADGLMAERYELLA 1735
DB 1141 GVFRQGCIAFRVITNNIDEEASNMEDVGMQDVHFNEVDYLMELLBOCADGLMAERYELLA 1200
QY 1736 DIYKLIIPYERKRDPERLAHLDTLHRAYSKYTEVMSGRLLAGTFRVAFFGQAQOYQ 1795
DB 1201 DIYKLIIPYERKRDPERLAHLDTLHRAYSKYTEVMSGRLLAGTFRVAFFGQAQOYQ 1260
QY 1796 FTDSEITDVBGFPEDEDGKEYIYKEKLTPLSISQRLKLVSDEKGSBNVKKIIDSQKN 1855
DB 1261 FTDSEITDVBGFPEDEDGKEYIYKEKLTPLSISQRLKLVSDEKGSBNVKKIIDSQKN 1320
QY 1856 PKDLSKAYIQVTHVIPFPDEKELQERKTEPERSHNIRPFEMPFTOTGRQGVBEQ 1915
DB 1321 PKDLSKAYIQVTHVIPFPDEKELQERKTEPERSHNIRPFEMPFTOTGRQGVBEQ 1380
QY 1916 CKRRTILTAHCFPYVKRIPVWYQHTDNLPIEVALIDEMSKVAELRQLCSSAEVDMIK 1975
DB 1381 CKRRTILTAHCFPYVKRIPVWYQHTDNLPIEVALIDEMSKVAELRQLCSSAEVDMIK 1440
QY 1976 LQKIQGSVSVVNNAGPLAIPADDTNTKYPNNKYKLKEVROPYACGALAVNE 2035
DB 1441 LQKIQGSVSVVNNAGPLAIPADDTNTKYPNNKYKLKEVROPYACGALAVNE 1500
QY 2036 RLKEDOLEYOEMKANTREMAKELSEIMHEOI 2068
DB 1501 RLKEDOLEYOEMKANTREMAKELSEIMHEOI 1533

RESULT 8
AAB36527
ID AAB36527 standard; Protein; 1352 AA.
XX AC AAB36527;
XX DT 07-MAR-2001 (first entry)
XX DE Human CLASP-2 protein sequence.
XX DB
XX KW CLASP-1; CLASP-2; transmembrane protein; immune response; inflammatory;
KW cadherin-like asymmetry protein; autoimmune disease; immunosuppressive;
KW immunomodulatory; antiinflammatory; antirheumatic; cycostatic;
KW hypotensive; antirheumatic; antianemic; haemostatic; neuroprotective;
KW hypersensitivity; transplantation rejection response; immunodeficiency;
KW proliferation; differentiation; inflammatory response; arthritis;
KW inflammatory bowel disease; haematopoietic cell; blood protein disorder;
KW anaemia; thrombocytopaenia; multiple sclerosis; rheumatoid arthritis;
KW endometriosis; pregnancy induced hypertension.
XX KW
XX OS Homo sapiens.
XX PN
XX PD WO200061747-A2.
XX PR 19-OCT-2000.
XX PF 13-APR-2000; 2000MO-US10158.
XX PR 14-APR-1999; 99US-0129171.
XX PR 14-MAY-1999; 99US-0134114.
XX PR 14-MAY-1999; 99US-0134117.
XX PR 14-MAY-1999; 99US-0134118.
XX PR 21-OCT-1999; 99US-0160860.
XX PR 29-OCT-1999; 99US-0162498.
XX PR 13-DEC-1999; 99US-0170453.
XX PR 14-JAN-2000; 2000US-0176195.

```


07-MAR-2001 (first entry)
 Human CLASP-2A protein sequence.

CLASP-1; CLASP-2; transmembrane protein; immune response; inflammatory; cadherin-like asymmetry protein; autoimmune disease; immunosuppressive; immunomodulatory; antiinflammatory; antirheumatic; cyostatic; hypotensive; antirheumatic; antianemic; haemostatic; neuroprotective; hypersensitivity; transplantation rejection response; immunodeficiency; proliferation; differentiation; inflammatory response; arthritis; inflammatory bowel disease; haematopoietic cell; blood protein disorder; anaemia; thrombocytopenia; multiple sclerosis; rheumatoid arthritis; endometriosis; pregnancy induced hypertension.

Homo sapiens.
 WO200061747-A2.

19-OCT-2000.
 13-APR-2000; 2000WO-US10158.

14-APR-1999; 99US-0129171.
 14-MAY-1999; 99US-0134114.
 14-MAY-1999; 99US-0134117.
 14-MAY-1999; 99US-0134118.
 21-OCT-1999; 99US-0160860.
 29-OCT-1999; 99US-0162498.
 13-DEC-1999; 99US-0170453.
 14-JAN-2000; 2000US-0176195.
 14-FEB-2000; 2000US-0182296.

(ARBO-) ARBOR VITA CORP.
 Lu PS;
 WPI; 2000-619230/59.
 N-PSDB; AAC87973.

Isolated cadherin-like asymmetry protein-2 polynucleotide and polypeptide used to diagnose, treat and prevent autoimmune diseases and inflammatory responses -

Example 1; Fig 2B; 286bp; English.

The present invention describes cadherin-like asymmetry protein-2 (CLASP-2). CLASP-2 can have immunosuppressive, immunomodulatory, antiinflammatory, antirheumatic, cyostatic, hypotensive, antirheumatic, antianemic, haemostatic and neuroprotective activities. CLASP-2 can be used to inhibit an immune response in a subject by interfering with the ability of a CLASP-2 protein to bind to another T cell or B cell. An immune response in a subject may also be inhibited by administering an antibody which specifically binds to CLASP-2. CLASP-2 polynucleotides, proteins and antibodies can be used to prevent or treat a CLASP-2 mediated disease, such as an autoimmune disease caused or exacerbated by increased activity of TH1 cells. They can also be used to treat hypersensitivities, prevent transplantation rejection responses and augment immune responsiveness in immunodeficiency states, inhibit proliferation and differentiation of cells involved in an inflammatory response e.g. arthritis, inflammatory bowel disease and increase differentiation and proliferation of haematopoietic cells e.g. to treat anaemia, thrombocytopenia and other blood protein disorders. Disorders treated by disrupting CLASP-2 function include multiple sclerosis, rheumatoid arthritis, endometriosis and pregnancy induced hypertension. The present sequence represents human CLASP-2A, which is used in the exemplification of the present invention.

Sequence 1352 AA;

Query Match 62.0%; Score 6784.5; DB 21; Length 1352;
 Best Local Similarity 92.7%; Pred. No. 0;
 Matches 1325; Conservative 1; Mismatches 3; Indels 101; Gaps 3;

QY 701 VLHHHNPDEYDEIKIELPTQLHEKHLLTFPHVSCNNSKSGTKRKDDVETQVGSWL 760
 DB 1 VLHHHNPDEYDEIKIELPTQLHEKHLLTFPHVSCNNSKSGTKRKDDVETQVGSWL 60
 QY 761 PLKDGRRVTSEOHIPVSNIPSGTIGQELGMRHVPETIKMDGGRPLKISHYVST 820
 DB 61 PLKDGRRVTSEOHIPVSNIPSGTIGQELGMRHVPETIKMDGGRPLKISHYVST 120
 QY 821 VYTDQHLNHFQYCKQKESGAQALGNELVYKLSLHMEGHVMAIFPTLLNQLFRVLT 880
 DB 121 VYTDQHLNHFQYCKQKESGAQALGNELVYKLSLHMEGHVMAIFPTLLNQLFRVLT 180
 QY 881 RATOEEVAVNTRVLIHVAAQCHEGBESHLRSYKVAKYKABPYVASYKTVHEELTKSM 940
 DB 181 RATOEEVAVNTRVLIHVAAQCHEGBESHLRSYKVAKYKABPYVASYKTVHEELTKSM 240
 QY 941 TTILKPSADPLTSNKLTKSWFFPDVLIKMAQHLIENSKYKTLRNGCFPPASVYHAAETV 1000
 DB 241 TTILKPSADPLTSNKLTKSWFFPDVLIKMAQHLIENSKYKTLRNGCFPPASVYHAAETV 300
 QY 1001 VNMLPHITTOKFRDNPESKSNANSLAVFIKRCFTFMDRGVFKOINNYISCFAPGDPKT 1060
 DB 301 VNMLPHITTOKFRDNPESKSNANSLAVFIKRCFTFMDRGVFKOINNYISCFAPGDPKT 360
 QY 1061 LFEYKFEFLRVVCSNHHYIPLNLPMPFGKGIQRYODQLDYSLTDFECRNHFLVGLLLR 1120
 DB 361 LFEYKFEFLRVVCSNHHYIPLNLPMPFGKGIQRYODQLDYSLTDFECRNHFLVGLLLR 420
 QY 1121 EVGTALQFREVRLAIVLKNLILKHSFDDRYASRSHQAATATLYLPLFGILLIENVORI 1180
 DB 421 EVGTALQFREVRLAIVLKNLILKHSFDDRYASRSHQAATATLYLPLFGILLIENVORI 480
 QY 1181 NVRDVSPPVNAAGMTVKDESIALPAVNPLVTPQKGSTLDNSLHKDLAIGIASPYTTS 1240
 DB 481 NVRDVSPPVNAAGMTVKDESIALPAVNPLVTPQKGSTLDNSLHKDLAIGIASPYTTS 540
 QY 1241 TPININSVANADSRGSLISTDSGNSLPERNSEKSNLDKHOQSGTLGNSVVRCDXLDQSEI 1300
 DB 541 TPININSVANADSRGSLISTDSGNSLPERNSEKSNLDKHOQSGTLGNSVVRCDXLDQSEI 600
 QY 1301 KSLMCFIYILKNSDDALFTYWNKASTSELMDFPTISEVCLHQYQWCKRYIA----- 1354
 DB 601 KSLMCFIYILKNSDDALFTYWNKASTSELMDFPTISEVCLHQYQWCKRYIA----- 660
 QY 1355 -----RTGMMHARLQOLGSLDNLSTFNHSYXGSHDADVLHQSLEANIA 1397
 DB 661 GPIYHDRKSQTLPIVSRNRTGMMHARLQOLGSLDNLSTFNHSYXGSHDADVLHQSLEANIA 720
 QY 1398 TEVCUTALDTLSLFTLAFKNQLDLADHGNPLMKKVFVYVLCFLQKQSEETALKVFTALR 1457
 DB 721 TEVCUTALDTLSLFTLAFKNQLDLADHGNPLMKKVFVYVLCFLQKQSEETALKVFTALR 780
 QY 1458 SLIYFPESTFYEGRADWCALCYEILKCCNSKLSIRTEASQLLYFLMRNPFDTYKKSIF 1517
 DB 781 SLIYFPESTFYEGRADWCALCYEILKCCNSKLSIRTEASQLLYFLMRNPFDTYKKSIF 840
 QY 1518 VRTHQVITISQILADVVGIGTRFOOSLIINNCAASDLIKHTSSSDVKDLTKIR 1577
 DB 841 VRTHQVITISQILADVVGIGTRFOOSLIINNCAASDLIKHTSSSDVKDLTKIR 900
 QY 1578 TVLMATAQMKHENDPEMLVDLOYSIAKSVASTBELRTKTLWDSMARIHVKNGDLSAAMC 1637
 DB 901 TVLMATAQMKHENDPEMLVDLOYSIAKSVASTBELRTKTLWDSMARIHVKNGDLSAAMC 960
 QY 1638 YVHTATVAEYLTREKAVQWEPPLIPSHSACLRRSGCVFRQGTAFRVITPINIDEAS 1697
 DB 961 YVHTATVAEYLTREKAVQWEPPLIPSHSACLRRSGCVFRQGTAFRVITPINIDEAS 997
 QY 1698 MMEDEVGMQDVHPNEDVLMELLECAGDLMKABRYELADIYKLIIFETEKRRDPERLNL 1757
 DB 998 MMEDEVGMQDVHPNEDVLMELLECAGDLMKABRYELADIYKLIIFETEKRRD----- 1051
 QY 1758 YDTLHRAYSKYTEVHSGRLLGTYFRVAFPGOAAOYQFTDSETDVEGFPEDBDGKEYIY 1817

```

Db      1052 -----|
Qy      1818 KEPKLPLSEISORLLKLYSDKFGSENVMMKIDSGKVNPKDLSKXAYIQTTHVLPFPE 1877
Db      1063 KEKPLPLSEISORLLKLYSDKFGSENVMMKIDSGKVNPKDLSKXAYIQTTHVLPFPE 1122
Qy      1878 KELOEKTEFERSHNRIRFMFEMPFTQTGRQGVVEOCKRRTILTAHCFPVKKRIV 1937
Db      1123 KELOEKTEFERSHNRIRFMFEMPFTQTGRQGVVEOCKRRTILTAHCFPVKKRIV 1182
Qy      1938 MYOHTDLNPIEVAIDEMSKKVAELPOLCSSAEVDMIKIQLKQGSVSVQVNAGPLAVAR 1997
Db      1183 MYOHTDLNPIEVAIDEMSKKVAELPOLCSSAEVDMIKIQLKQGSVSVQVNAGPLAVAR 1242
Qy      1998 AFLDDNTKRYPNKVKLLKEVROFEVACGALANBERLIKEDOLEYOEEMKANVREMA 2057
Db      1243 AFLDDNTKRYPNKVKLLKEVROFEVACGALANBERLIKEDOLEYOEEMKANVREMA 1302
Qy      2058 KEISEIMHBOICPLEBKTSLVPLNSLHIFNAISGTPSTVHWGHTSSSVV 2107
Db      1303 KEISEIMHBOICPLEBKTSLVPLNSLHIFNAISGTPSTVHWGHTSSSVV 1352

```

```

RESULT 10
ABG61670
ID ABG61670 standard; Protein; 1353 AA.
XX
AC ABG61670;
XX
DT 13-AUG-2002 (first entry)
XX
DE Cadherin-like asymmetry protein (CLASP).
XX
KW Human; autoimmune diseases; haematopoietic disorder; Digorge syndrome;
KW blood protein disorder; agammaglobulinemia; dysgammaglobulinemia;
KW ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
KW thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
KW haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;
KW endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
KW autoimmune pulmonary inflammation; organ rejection; inflammation;
KW CLASP.
XX
OS Homo sapiens.
XX
PN MO20023117-A2.
XX
PD 18-APR-2002.
XX
PF 15-OCT-2001; 2001WO-US32202.
XX
PR 13-OCT-2000; 2000US-0687837.
XX
PA (ARBO-) ARBOR VITA CORP.
PA (GARM/) GARMAN J D.
PA (CAND/) CANDIA A F.
XX
PI Lu PS;
XX
DR WPI: 2002-416861/44.
DR N-PSDB: ABR84964.
XX
PT New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
PT an immune response, and for treating multiple sclerosis, rheumatoid
PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
PT and sepsis
XX
PS Disclosure; Figure 1; 245pp; English.
XX
CC The invention relates to an isolated polypeptide (I) comprising an amino
CC acid sequence that has 90 % sequence identity to one of the human
CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E)
CC sequences (PS). (I) is useful for identifying a compound or agent that

```

CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for
 CC inhibiting an immune response in a subject. A pharmaceutical composition
 CC comprising a nucleic acid encoding (I), or (II) is useful for preventing
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
 CC the autoimmune disease is caused or exacerbated by increased activity
 CC of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2
 CC polynucleotide or fragments can be used in diagnostics (e.g., as probes
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2
 CC polynucleotides can screen for CLASP-2 agonists and antagonists.
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or
 CC disorders of the immune system, by activating or inhibiting the
 CC activation, differentiation of immune cells and can treat or detect
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
 CC or polynucleotides can increase differentiation and proliferation of
 CC haematopoietic cells, including the pluripotent stem cells to treat those
 CC disorders associated with a decrease in certain (or many) types of
 CC haematopoietic cells e.g., immunologic deficiency syndromes including
 CC blood protein disorders (e.g., agammaglobulinemia,
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable
 CC immunodeficiency, Digorge syndrome, lymphopenia, thrombocytopenia, or
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or
 CC hypersensitivity to an antigenic molecules, organ rejection or graft-
 CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61708
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the
 CC invention.

CC Sequence 1353 AA;

Query Match 62.0%; Score 6784.5; DB 23; Length 1353;
 Best Local Similarity 92.7%; Pred. No. 0;
 Matches 1325; Conservative 1; Mismatches 3; Indels 101; Gaps 3;

```

Qy      701 VLHHNQNEPFYDEIKELPTQLHEKHLLTFHFVSCNNSKSGSKKDVETQVGYGWL 760
Db      1 VLHHNQNEPFYDEIKELPTQLHEKHLLTFHFVSCNNSKSGSKKDVETQVGYGWL 60
Qy      761 PLKDGRRVTSBOHIVSANTPSGYLGQYELGMRHGYPEIKWWDGKPLKISTHLYST 820
Db      61 PLKDGRRVTSBOHIVSANTPSGYLGQYELGMRHGYPEIKWWDGKPLKISTHLYST 120
Qy      821 VYTDOPHLNPFQYCKQTESGAQAGNELVYVYLSLHMEGHVMIAPLTLNOLFRLVT 880
Db      121 VYTDOPHLNPFQYCKQTESGAQAGNELVYVYLSLHMEGHVMIAPLTLNOLFRLVT 180
Qy      881 RATOEEVAVNTTRVLIHVVAQCHEGLESLSRYVYKVAKEPYVASEYKTVHEELTSM 940
Db      181 RATOEEVAVNTTRVLIHVVAQCHEGLESLSRYVYKVAKEPYVASEYKTVHEELTSM 240
Qy      941 TTILKPSADPLTSNKLKTYSPFPDVLTKSMAOHLIENSKYKLLRNQFPASVYHNAVETV 1000
Db      241 TTILKPSADPLTSNKLKTYSPFPDVLTKSMAOHLIENSKYKLLRNQFPASVYHNAVETV 300
Qy      1001 VNMMPHITQKFRPNPKASKNANSLAVIRGCTFMDRGVFPQINNYISGPAQDPKT 1060
Db      301 VNMMPHITQKFRPNPKASKNANSLAVIRGCTFMDRGVFPQINNYISGPAQDPKT 360
Qy      1061 LFEYKFEFLRVVCHNEHTIPLNLPMPFGKRIQRYQDQLDYSLTDECRNHFVLGLLR 1120
Db      361 LFEYKFEFLRVVCHNEHTIPLNLPMPFGKRIQRYQDQLDYSLTDECRNHFVLGLLR 420
Qy      1121 EVGTALQSFREVRLLAISVLNLLIKHSFDRVYASRSHQARIATLYLPFGILLIENVORI 1180

```

```

Db      421 EVGIALQEFREVRLIAISVLKNLLIKHSFDRYVARSHQARIATITLPLFGLLIENVORI 480
Qy      1181 NVRDVPFPVNAAGTVKDESLAPAVNPVLPVPOKSTLDNSLHKDLGASIGIAPYTT 1240
Db      481 NVRDVPFPVNAAGTVKDESLAPAVNPVLPVPOKSTLDNSLHKDLGASIGIAPYTT 540
Qy      1241 TPINISVRNADSRGSLISTDSGNSLPERNSEKNSLDKHCQSSITGNSVVRCDLQDSEI 1300
Db      541 TPINISVRNADSRGSLISTDSGNSLPERNSEKNSLDKHCQSSITGNSVVRCDLQDSEI 600
Qy      1301 KSLMLCFYIILKMSDDALFTYWNKASTSELMDFPITISVCLHQOYMGKRYIA----- 1354
Db      601 KSLMLCFYIILKMSDDALFTYWNKASTSELMDFPITISVCLHQOYMGKRYIA----- 1354
Qy      1355 -----RTGMHARLQOLGSLDNLTFNHSYGSHDADVLHOSLEANIA 1397
Db      661 GPVYDRKSQTLPSGRNRTGMHARLQOLGSLDNLTFNHSYGSHDADVLHOSLEANIA 1457
Qy      1398 TEVCJTALDTSLSFTLAFKNQDLADHGNPLMKKVFDDVYLCFLQKHQSEFALKNVFTALR 1457
Db      721 TEVCJTALDTSLSFTLAFKNQDLADHGNPLMKKVFDDVYLCFLQKHQSEFALKNVFTALR 1457
Qy      1458 SLIYFPSTFYEGRADMCALCYEILKCNKSLSSITRTASQLIFLWNNFDYTGKSP 1517
Db      781 SLIYFPSTFYEGRADMCALCYEILKCNKSLSSITRTASQLIFLWNNFDYTGKSP 1517
Qy      1518 VTHLQVITISVSQLADVVIGIGTRFOOSLTIINNCANSDRILKHTSPSSDYKDLTKRIR 1577
Db      841 VTHLQVITISVSQLADVVIGIGTRFOOSLTIINNCANSDRILKHTSPSSDYKDLTKRIR 1577
Qy      1578 TVLMAVTAQMKHEHNDPEMLVDLOYSLAKSVASTPELRTKWTMDSMARHIVKNDLSEAMC 1637
Db      901 TVLMAVTAQMKHEHNDPEMLVDLOYSLAKSVASTPELRTKWTMDSMARHIVKNDLSEAMC 1637
Qy      1638 YVHTVTAVALAEYILTRKBAVOMEPPLPHSHSACLRSSRGCVFPOGCTAFVITPNIDEAS 1697
Db      961 YVHTVTAVALAEYILTRKBAVOMEPPLPHSHSACLRSSRGCVFPOGCTAFVITPNIDEAS 1697
Qy      1698 NMEDVGMODVHFNEDVLMELBEOCADGLMKARVELIADYLIPIYKRRDFERLAML 1757
Db      998 NMEDVGMODVHFNEDVLMELBEOCADGLMKARVELIADYLIPIYKRRDFERLAML 1757
Qy      1758 YDTLHRAVSKTVBVMHSGRRLLGTFRVAFPGQAQYQITDSEITVEGFEDEDEGKEYIY 1817
Db      1052 -----FEDEDEGKEYIY 1062
Qy      1818 KEPKLTPLUSEISQRLIKLYSDKRGSENVMIODSGKVNPKDLSKYAYIQVTHVIPFDE 1877
Db      1063 KEPKLTPLUSEISQRLIKLYSDKRGSENVMIODSGKVNPKDLSKYAYIQVTHVIPFDE 1877
Qy      1878 KELORKTEFEBSHNIRRFMPMPPTQNGKRGVBEQCKRRTILTAHCFPVYKKRIPV 1937
Db      1123 KELORKTEFEBSHNIRRFMPMPPTQNGKRGVBEQCKRRTILTAHCFPVYKKRIPV 1937
Qy      1938 MYQHTDNLPIEVAIDEMSKVAELRQLCSSAEVMIKQLKQSSVSQVWAGPLAYAR 1997
Db      1183 MYQHTDNLPIEVAIDEMSKVAELRQLCSSAEVMIKQLKQSSVSQVWAGPLAYAR 1997
Qy      1998 AFLDQNTTKRYPDNKKYLKEYVPROFVEACGALLAVNERLIEDOLEYOEMKANYREMA 2057
Db      1243 AFLDQNTTKRYPDNKKYLKEYVPROFVEACGALLAVNERLIEDOLEYOEMKANYREMA 2057
Qy      2058 KELSIMEHOICPLEEKTSLVNSLIHFAISGTPSTMVHGMTSSSVV 2107
Db      1303 KELSIMEHOICPLEEKTSLVNSLIHFAISGTPSTMVHGMTSSSVV 2107

```

```

RESULT 11
ABG61671
ID ABG61671 standard; Protein; 1353 AA.
AC ABG61671;
XX
XX

```

```

DT      13-AUG-2002 (first entry)
DE      Cadherin-like asymmetry protein (CLASP) isoform.
XX
XX
KM      Human; autoimmune disease; haematopoietic disorder; Digorge syndrome;
KM      blood protein disorder; agammaglobulinaemia; dysgammaglobulinaemia;
KM      ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
KM      thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
KM      haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;
KM      endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
KM      autoimmune pulmonary inflammation; organ rejection; inflammation;
KM      CLASP.
XX
OS      Homo sapiens.
XX
PN      WO2002/1117-A2.
XX
PD      18-APR-2002.
XX
PF      15-OCT-2001; 2001WO-US22202.
XX
PR      13-OCT-2000; 2000US-0687837.
XX
PA      (ARBO-) ARBOR VITA CORP.
XX
PA      (GARM/) GARMAN J D.
XX
PA      (CAND/) CANDIA A F.
XX
PI      Lu PS;
XX
DR      WPI; 2002-416861/44.
XX
DR      N-PSDB; ABK84965.
XX
PT      New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
PT      an immune response, and for treating multiple sclerosis, rheumatoid
PT      arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
PT      and sepsis
XX
XX
PS      Disclosure; Figure 2; 245pp; English.
XX
XX
CC      The invention relates to an isolated polypeptide (I) comprising an amino
CC      acid sequence that has 90 % sequence identity to one of the human
CC      cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2a, 2b, 2c, 2e)
CC      sequences (PS). (I) is useful for identifying a compound or agent that
CC      binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
CC      detecting a CLASP-2 polypeptide in a sample. (II) is useful for
CC      inhibiting a immune response in a subject. A pharmaceutical composition
CC      comprising a nucleic acid encoding (I), or (II) is useful for preventing
CC      or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
CC      the autoimmune disease is caused or exacerbated by increased activity
CC      of Th1 cells. CLASP-2 polynucleotides are useful as probes or primers for
CC      detection or inhibition of CLASP-2 expression (e.g., antisense or
CC      ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
CC      polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
CC      antibodies or are used as therapeutic polypeptides. The CLASP-2
CC      polynucleotide or fragments can be used in diagnostics (e.g., as probes
CC      for CLASP-2 expression), as a lymphocyte marker and for therapeutic
CC      purposes. CLASP-2 polynucleotides can construct transgenic and knockout
CC      animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2
CC      polynucleotides can screen for CLASP-2 agonists and antagonists.
CC      CLASP-2 polypeptides or polynucleotides can treat deficiencies or
CC      disorders of the immune system, by activating or inhibiting the
CC      activation, differentiation of immune cells and can treat or detect
CC      deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
CC      or polynucleotides can increase differentiation and proliferation of
CC      haematopoietic cells, including the pluripotent stem cells to treat those
CC      disorders associated with a decrease in certain (or many) types of
CC      haematopoietic cells e.g., immunologic deficiency syndromes including
CC      blood protein disorders (e.g., agammaglobulinaemia,
CC      dysgammaglobulinaemia, ataxia telangiectasia, common variable
CC      immunodeficiency, Digorge syndrome, lymphopenia, thrombocytopenia, or
CC      haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
CC      detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
CC      Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,

```


CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or
 CC hypersensitivity to an antigenic molecule, organ rejection or graft-
 CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61708
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the
 CC invention.

XX
 SQ Sequence 1353 AA;

Query Match 62.0%; Score 6784.5; DB 23; Length 1353;
 Best Local Similarity 92.7%; Pred. No. 0;
 Matches 1353; Conservative 1; Mismatches 3; Indels 101; Gaps 3;

QY 701 VLHHNHNPFEYDEIKELPTQLHEKHHLLTFFHVSCHNSKSGSTKRDVETOVGSWL 760
 DB 1 VLHHNHNPFEYDEIKELPTQLHEKHHLLTFFHVSCHNSKSGSTKRDVETOVGSWL 60

QY 761 PLLKGRVVTSEQHIVSANLPSGYLGYOELGMRHGYBEIKAVDGGKPLKISTHLYST 820
 DB 61 PLLKGRVVTSEQHIVSANLPSGYLGYOELGMRHGYBEIKAVDGGKPLKISTHLYST 120

QY 821 VYTODHNLNFPYCOKTESGAQALGNELVKYLSLHAMEGHMIAFLPTILNQLFRVLT 880
 DB 121 VYTODHNLNFPYCOKTESGAQALGNELVKYLSLHAMEGHMIAFLPTILNQLFRVLT 180

QY 881 RATOEEVAVNVTREVIHVVAQCHEEGLESLSYKAVYKAEPPYVASEYKTVHEELTKSM 940
 DB 181 RATOEEVAVNVTREVIHVVAQCHEEGLESLSYKAVYKAEPPYVASEYKTVHEELTKSM 240

QY 941 TTLTKSADFLTSNKLKTSWFFDVLKISMAOHLIENSKVKLLRNQRPASVYHAAVEV 1000
 DB 241 TTLTKSADFLTSNKLKTSWFFDVLKISMAOHLIENSKVKLLRNQRPASVYHAAVEV 300

QY 1001 VNNLMHITQKPEFNDPEASKNANSLAVFKRCFTMDRGFVFKOINNYISCPAPDPT 1060
 DB 301 VNNLMHITQKPEFNDPEASKNANSLAVFKRCFTMDRGFVFKOINNYISCPAPDPT 360

QY 1061 LFEYKEEFLRVNCGNHHYIPLNLPMPFGKRIQRYODLDLDVSLTDFCRNHFVGLLR 1120
 DB 361 LFEYKEEFLRVNCGNHHYIPLNLPMPFGKRIQRYODLDLDVSLTDFCRNHFVGLLR 420

QY 1121 EVGTALQERREVRLLAISVKNLLIGSPDDRYSASHQARIATVLPFGLLIENVQRI 1180
 DB 421 EVGTALQERREVRLLAISVKNLLIGSPDDRYSASHQARIATVLPFGLLIENVQRI 480

QY 1181 NVRDVPFPVNAGMTVKDSLALPAINPLVTPQKGSITLNSLHKDLGASIGIASPYTS 1240
 DB 481 NVRDVPFPVNAGMTVKDSLALPAINPLVTPQKGSITLNSLHKDLGASIGIASPYTS 540

QY 1241 TPNINSVRNADSRGSLISTDSGNSLPERNSEKNSLDKIQOSSLTGNVVRCDKLQSEI 1300
 DB 541 TPNINSVRNADSRGSLISTDSGNSLPERNSEKNSLDKIQOSSLTGNVVRCDKLQSEI 600

QY 1301 KSLIMCFYILKMSMDALFTYWNKASTSELDFFITSVCLHOFQYMGKRYIARNOEGL 1354
 DB 601 KSLIMCFYILKMSMDALFTYWNKASTSELDFFITSVCLHOFQYMGKRYIARNOEGL 660

QY 1355 -----RTGMHARLOQLGSLDNLTPNHSYGHSHADVLHOSILEANTA 1397
 DB 661 GPTVHDRKQOTLPVSNKRTGMHARLOQLGSLDNLTPNHSYGHSHADVLHOSILEANTA 720

QY 1398 TEVCLTALDPLSLFTLAFKQQLADHGNPLMKVVDVYLCPLOKQSEETALKVVTALR 1457
 DB 721 TEVCLTALDPLSLFTLAFKQQLADHGNPLMKVVDVYLCPLOKQSEETALKVVTALR 780

QY 1458 SLIYKFPSTFYEGRADMCALCYEILKCCNSKLSIRTEASQLLYFLMRNPFDTGKSF 1517
 DB 781 SLIYKFPSTFYEGRADMCALCYEILKCCNSKLSIRTEASQLLYFLMRNPFDTGKSF 840

QY 1518 VRTHLQVITISVQSLADVIGIGTRFOQSLIINNCAANDRLIKHSPSSDVYDLTKRIR 1577
 DB 841 VRTHLQVITISVQSLADVIGIGTRFOQSLIINNCAANDRLIKHSPSSDVYDLTKRIR 900

QY 1578 TVLMATAQWKEHENDPEMLVDLOYSLAKSVASTPELRKTWLDNSNARIHVKNQDLSSEAMC 1637
 DB 901 TVLMATAQWKEHENDPEMLVDLOYSLAKSVASTPELRKTWLDNSNARIHVKNQDLSSEAMC 960

QY 1638 YVHTVTAALVAEYITREKAVQWEPPLPHSHSACLRSGVRRQGTARVVTNNIDEAS 1697
 DB 961 YVHTVTAALVAEYITREKAVQWEPPLPHSHSACLRSGVRRQGTARVVTNNIDEAS 997

QY 1698 NMEEDVGMQDVHFNEDVLMELLBOCADGLMKAEYELIADIYKLIPIYEKRDFFERLAML 1757
 DB 998 NMEEDVGMQDVHFNEDVLMELLBOCADGLMKAEYELIADIYKLIPIYEKRDFFERLAML 1051

QY 1758 YDTHRAVSKYTVNHSGRRLITGVFVAFPGQAQYQFTDSEITDVEGFEDDEGKEYIY 1817
 DB 1052 -----FEDEDEGKEYIY 1062

QY 1818 KEPKLTPLSEISQRLKLYSDKFGSENVKMTODSKVNPKDLDSKAYITQVTHVLPFDE 1877
 DB 1063 KEPKLTPLSEISQRLKLYSDKFGSENVKMTODSKVNPKDLDSKAYITQVTHVLPFDE 1122

QY 1878 KELQERKTEFERSHNIRRFMPFPTQTKQGGVEBOCKRTITLTAHCFPYVKGRIPV 1937
 DB 1123 KELQERKTEFERSHNIRRFMPFPTQTKQGGVEBOCKRTITLTAHCFPYVKGRIPV 1182

QY 1938 MYQHTDNLPIEVALIDENSKKVAELRQCSSAEVDMIKLOLQOGSVSVQVNAEGLAYAR 1997
 DB 1183 MYQHTDNLPIEVALIDENSKKVAELRQCSSAEVDMIKLOLQOGSVSVQVNAEGLAYAR 1242

QY 1998 AFLDITNTKRYPDNKVLLKEVFRQFVACGQALANERLLKEDOLEYOEEMKANYREMA 2057
 DB 1243 AFLDITNTKRYPDNKVLLKEVFRQFVACGQALANERLLKEDOLEYOEEMKANYREMA 1302

QY 2058 KELSEIMEOICPLEEKTSTVLPNSLHIFNAISGTPSTVMVGMSTSSSV 2107
 DB 1303 KELSEIMEOICPLEEKTSTVLPNSLHIFNAISGTPSTVMVGMSTSSSV 1352

RESULT 12
 ID ABG61672
 ID ABG61672 standard; Protein; 1353 AA.

XX ABG61672;
 XX
 DT 13-AUG-2002 (first entry)
 XX
 DE Cadherin-like asymmetry protein (CLASP) isoform #1.

XX Human; autoimmune disease; haematopoietic disorder; Digesorge syndrome;
 KW blood protein disorder; agammaglobulinemia; dysgammaglobulinemia;
 KW ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
 KW chromocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
 KW haemolytic anemia; multiple sclerosis; rheumatoid arthritis; lupus;
 KW endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
 KW autoimmune pulmonary inflammation; organ rejection; inflammation;
 KW CLASP.

OS Homo sapiens.
 XX
 PN WO200231117-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 15-OCT-2001; 2001WO-US32202.
 XX
 PR 13-OCT-2000; 2000US-0687837.
 XX
 PA (ARBO-) ARBOR VITA CORP.
 PA (GARM/) GARMAN J D.
 PA (CAND/) CANDIA A F.
 XX
 PI Lu PS;
 XX
 DR WPI; 2002-416861/44.

DR N-PSDB; ABK84966.

PT New human cathenin-1-like asymmetry proteins (CLASP)-2 for modulating
PT an immune response, and for treating multiple sclerosis, rheumatoid
PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
PT and sepsis -

PS Disclosure; Figure 3B; 245pp; English.

The invention relates to an isolated polypeptide (I) comprising an amino acid sequence that has 90 % sequence identity to one of the human cadherin-1-like asymmetry proteins (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E) sequences (BS). (I) is useful for identifying a compound or agent that binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for detecting a CLASP-2 polypeptide in a sample. (II) is useful for inhibiting a immune response in a subject. A pharmaceutical composition comprising a nucleic acid encoding (I), or (II) is useful for preventing or treating a CLASP-2 mediated disease e.g., an autoimmune disease, where the autoimmune disease is caused or exacerbated by increased activity of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for detection or inhibition of CLASP-2 expression (e.g., antisense or ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2 polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-antibodies or are used as therapeutic polypeptides. The CLASP-2 polynucleotide or fragments can be used in diagnostics (e.g., as probes for CLASP-2 expression), as a lymphocyte marker and for therapeutic purposes. CLASP-2 polynucleotides can construct transgenic and knockout animals, e.g., for screening of CLASP-2 agonists and antagonists. polynucleotides can screen for CLASP-2 agonists and antagonists. CLASP-2 polypeptides or polynucleotides can treat deficiencies or disorders of the immune system, by activating or inhibiting the activation, differentiation of immune cells and can treat or detect deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides or polynucleotides can increase differentiation and proliferation of haematopoietic cells, including the pluripotent stem cells to treat those disorders associated with a decrease in certain (or many) types of haematopoietic cells e.g., immunologic deficiency syndromes including blood protein disorders (e.g., agammaglobulinemia, dysgammaglobulinemia, ataxia telangiectasia, common variable immunodeficiency, Disgorge syndrome, lymphopenia, thrombocytopenia, or haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia, Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus, endometriosis, autoimmune thyroiditis, and autoimmune pulmonary inflammation. CLASP-2 can be used to treat organopathy or hyperreactivity to an antigenic molecules, organ rejection or graft-versus-host disease (GVHD) and inflammation. ABB51670-AB651708 represent cadherin-1-like asymmetry protein (CLASP) sequences of the invention.

SQ Sequence 1353 AA;

Query Match	Score	DB	Length
62.0%	6784.5	23	1353

Matches 1325; Conservative 1; Mismatch

Matches 1325; Conservative 1; Mismatches 3; Indels 101; Gaps 3

Qy	701	VLHHQNEPFADEIKETLPOLHEGHHLLTFPHNSCNSSGSKGKDDVETQVGYM.L	760
Db	1	VLHHQNEPFADEIKETLPOLHEGHHLLTFPHNSCNSSGSKGKDDVETQVGYM.L	60
Qy	761	PLLDGRVVTSEOHIPVANSIPSGYLGYOELGMGNGHCPLEIKWDGKPLLKISTH.LVST	820
Db	61	PLLDGRVVTSEOHIPVANSIPSGYLGYOELGMGNGHCPLEIKWDGKPLLKISTH.LVST	120
Qy	821	VYTDQOH.LHNFQYCKTESGAQALGNELVYKYLKSLHMEGVMYLAFLPTILNQLFR.LLT	880
Db	121	VYTDQOH.LHNFQYCKTESGAQALGNELVYKYLKSLHMEGVMYLAFLPTILNQLFR.LLT	180
Qy	881	RATQEEVAVNVTYRVIIIVVAQCHEEGESLHLSRYKVAUYKAPYAASEKTVHEBLTYSM	940
Db	181	RATQEEVAVNVTYRVIIIVVAQCHEEGESLHLSRYKVAUYKAPYAASEKTVHEBLTYSM	240
Qy	941	TTILKPSADFLTSNKLLKYSWFDFDVLIKSMAQHLIENSKVYKLNQSPFASVYNHAAVETV	1000

Db	241	TTTLKPSADPFLTSKULRYSWFFPDVILKSMAGLILNSKYKILRNQFRPASYHHAETV	300
Qy	1001	VNMMLPHITQKFRDNPEASKNANSLAVFIKRCFTFMDRGVFVKOINNYISCFAPGDKT	1060
Db	301	VNMMLPHITQKFGDNPEASKNANSLAVFIKRCFTFMDRGVFVKOINNYISCFAPGDKT	360
Qy	1061	LFEVKEPEFLRVVCHHEHYITPLNPMRGKGRIGORYOLOLDYSITDEPCRNHPLVGLLR	1120
Db	361	LFEVKEPEFLRVVCHHEHYITPLNPMRGKGRIGORYOLOLDYSITDEPCRNHPLVGLLR	420
Qy	1121	EVGVALDEFEVRRLIASVYKNLILKISPDORVSRSHQRIATLYPLFGLLIENORI	1180
Db	421	EVGVALDEFEVRRLIASVYKNLILKISPDORVSRSHQRIATLYPLFGLLIENORI	480
Qy	1181	NVRDVSFPVNAGMTVDESILAPVNPVLYTPQKSTJLDNSLHKDLIGALISGAPYTS	1240
Db	481	NVRDVSFPVNAGMTVDESILAPVNPVLYTPQKSTJLDNSLHKDLIGALISGAPYTS	540
Qy	1241	TPNINSVRNADSRGLISTDSGNSLPERNSEKNSLDKQOOSTJGNSVVRCDKLDQSEI	1300
Db	541	TPNINSVRNADSRGLISTDSGNSLPERNSEKNSLDKQOOSTJGNSVVRCDKLDQSEI	600
Qy	1301	KSILMCELYILIKMSD DALFTYNNKASTSELMDEFTISEVCLHOFQYMGKRYIA-----	1354
Db	601	KSILMCELYILIKMSD DALFTYNNKASTSELMDEFTISEVCLHOFQYMGKRYIARNQBL	660
Qy	1355	-----RTGMMHARLQOLSLDLSLFPNNSYGHSDADVLHQSLLEANIA	1397
Db	661	GPIYHDKSQCLPVSRRRTGMHARLQOLSLDLSLFPNNSYGHSDADVLHQSLLEANIA	720
Qy	1398	TEVCLTALDLSLFTLAFKQQLADHGNPLMKKVPDVLVCLPQKHQSETLAKNVPALR	1457
Db	721	TEVCLTALDLSLFTLAFKQQLADHGNPLMKKVPDVLVCLPQKHQSETLAKNVPALR	780
Qy	1458	SLIYKEPSTPEEGADMCALCYEILCCNSKJLSITTEASQULYELMRNPDVTYGGKSF	1517
Db	781	SLIYKEPSTPEEGADMCALCYEILCCNSKJLSITTEASQULYELMRNPDVTYGGKSF	840
Qy	1518	VRTHLQYIIVSOLIAOVGIGGRFOOSLSIINNCA NSRLIGHTFSSDPVOKLTRIR	1577
Db	841	VRTHLQYIIVSOLIAOVGIGGRFOOSLSIINNCA NSRLIGHTFSSDPVOKLTRIR	900
Qy	1578	TVLMAIQAOMKEHNDPEMLVDLOYSLAKSYASTBELKTYLMDSMARIHVXGDLSEAMC	1637
Db	901	TVLMAIQAOMKEHNDPEMLVDLOYSLAKSYASTBELKTYLMDSMARIHVXGDLSEAMC	960
Qy	1638	YVHTALVABEILTRKEAVQMBPRLPHSHACLRSHSGVFPROCTAPRYITPNIIDEAS	1697
Db	961	YVHTALVABEILTRK-----GVFRQCTAFRYITPNIIDEAS	997
Qy	1698	MMEBVGQDVHPFNDVAMELLEOCADQLMKAERELIADYIKLIIPIYEKRDDEBLAHL	1757
Db	998	MMEBVGQDVHPFNDVAMELLEOCADQLMKAERYELIADYIKLIIPIYEKRD-----	1051
Qy	1758	YDTLHRAYSKVTBVMHSGRRLGTFRVVAFFGQAAYOFTDSETVDEGFEDDEGKEYIY	1817
Db	1052	-----FEDEGKEYIY	1062
Qy	1818	KEPCLTJLSEISQRLKLYSDKRGSENVKMIODSGKNPKDLDISKYIYQVTHVIPFDE	1877
Db	1063	KEPCLTJLSEISQRLKLYSDKRGSENVKMIODSGKNPKDLDISKYIYQVTHVIPFDE	1122
Qy	1878	KELOERTEFERBSNINIRPFMEPFOTGRQGVESQCRRTITLAIHCPYVKKIIPV	1937
Db	1123	KELOERTEFERBSNINIRPFMEPFOTGRQGVESQCRRTITLAIHCPYVKKIIPV	1182
Qy	1938	MYQHHTDLNPIEVAIDEMSKKVAELROLCSASAEVDMIKLOKLGYSVVOVNAGPLAVAR	1997
Db	1183	MYQHHTDLNPIEVAIDEMSKKVAELROLCSASAEVDMIKLOKLGYSVVOVNAGPLAVAR	1242
Qy	1998	AFUDDNITKRYPDNKVTLKEVFRQVEACGQALAVNERLIKEDQLEYOEBMKANYREMA	2057

Db 1243 AFLDDNTKRYYPDNKVKLLKEVERQVEACGQALAVNERLIKEDQLEVOEMKANRYENA 1302
 QY 2058 KELSEIMHBOICPLEEKTSVLPNSLHIFNAISGTPSTWGHGTSSSVY 2107
 Db 1303 KELSEIMHBOICPLEEKTSVLPNSLHIFNAISGTPSTWGHGTSSSVY 1352

RESULT 13
 ABG61686
 ABG61686 standard; Protein; 1353 AA.
 ABG61686;
 13-AUG-2002 (first entry)
 Cadherin-like asymmetry protein (CLASP) isoform.
 Human; autoimmune disease; haematopoietic disorder; Digeorge syndrome;
 blood protein disorder; agammaglobulinemia; dysgammaglobulinemia;
 ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
 thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
 haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;
 endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
 autoimmune pulmonary inflammation; organ rejection; inflammation;
 CLASP.
 Homo sapiens.
 MO200231117-A2.
 18-APR-2002.
 15-OCT-2001; 2001MO-US32202.
 13-OCT-2000; 2000US-0687837.
 (ARBO-) ARBOR VITA CORP.
 (GARM/) GARMAN J D.
 (CAND/) CANDIA A F.
 Lu PS;
 WPI; 2002-416861/44.
 N-PSDB; ABK84973.
 New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
 an immune response, and for treating multiple sclerosis, rheumatoid
 arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
 and sepsis -
 Example 4; Figure 6A; 245p; English.

CC disorders of the immune system, by activating or inhibiting the
 CC activation, differentiation of immune cells and can treat or detect
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
 CC or polynucleotides can increase differentiation and proliferation of
 CC haematopoietic cells, including the pluripotent stem cells to treat those
 CC disorders associated with a decrease in certain (or many) types of
 CC haematopoietic cells e.g., immunologic deficiency syndromes including
 CC blood protein disorders (e.g., agammaglobulinemia,
 CC dysgammaglobulinemia, ataxia telangiectasia, common variable
 CC immunodeficiency, Digeorge syndrome, lymphopenia, thrombocytopenia, or
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or
 CC hypersensitivity to an antigenic molecules, organ rejection or graft-
 CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61708
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the
 CC invention.

Sequence 1353 AA;
 SQ
 Query Match 62.0%; Score 6784.5; DB 23; Length 1353;
 Best Local Similarity 92.7%; Pred. No. 0;
 Matches 1325; Conservative 1; Mismatches 3; Indels 101; Gaps 3;

QY 701 VLHHQNEPEYDEIKELPTOLHEKHHLLTFPHVSCNNSKSGTKRBDVETQVGSML 760
 Db 1 VLHHQNEPEYDEIKELPTOLHEKHHLLTFPHVSCNNSKSGTKRBDVETQVGSML 60
 QY 761 PLKDGRRVTSBOHIVPISANLPSGYLGYOELGMRHRYPEIKWVDGKPLKISTHLVST 820
 Db 61 PLKDGRRVTSBOHIVPISANLPSGYLGYOELGMRHRYPEIKWVDGKPLKISTHLVST 120
 QY 821 VYTOPQHLNPFQYQCKTESGQALGNELVYKSLHMEGHVMAFLPTLLNQFRVLT 880
 Db 121 VYTOPQHLNPFQYQCKTESGQALGNELVYKSLHMEGHVMAFLPTLLNQFRVLT 180
 QY 881 RATOEVAVNVTWVRIHVVAQCHEGLSHLSRYKYVYKAPYVASEKYVHEELTSM 940
 Db 181 RATOEVAVNVTWVRIHVVAQCHEGLSHLSRYKYVYKAPYVASEKYVHEELTSM 240
 QY 941 TTIKPSADFLTSNKLKYSWFFPDVLKSAQHLIENSXYKLLNQFPASVTHAVETV 1000
 Db 241 TTIKPSADFLTSNKLKYSWFFPDVLKSAQHLIENSXYKLLNQFPASVTHAVETV 300
 QY 1001 VNMMLPHITQKRPDPEASKANHSLAVFTRCPFFMRGVRPKINNYISCFAGDPT 1060
 Db 301 VNMMLPHITQKRPDPEASKANHSLAVFTRCPFFMRGVRPKINNYISCFAGDPT 360
 QY 1061 LPEYKFEFLRVVCMNEHYIPLNLPMPFGKGRIGRQDQLDYSLTDEFCRNHFLVGLLR 1120
 Db 361 LPEYKFEFLRVVCMNEHYIPLNLPMPFGKGRIGRQDQLDYSLTDEFCRNHFLVGLLR 420
 QY 1121 EVGTALQEFREVRLLAISVLKXLLKHSFDDRYASRSHQARIATLYLPFGLLIENVORI 1180
 Db 421 EVGTALQEFREVRLLAISVLKXLLKHSFDDRYASRSHQARIATLYLPFGLLIENVORI 480
 QY 1181 NVRDVSPPPVNAAGMTVKDESLAPVAVNPLVPQKSTLDNSLHMDLGAIGISAPYTT 1240
 Db 481 NVRDVSPPPVNAAGMTVKDESLAPVAVNPLVPQKSTLDNSLHMDLGAIGISAPYTT 540
 QY 1241 TPNIINSVNAASRGSLITDSGNSLPEKNSEKNSLDHGQOSTLGNSVBCDKLDQSEI 1300
 Db 541 TPNIINSVNAASRGSLITDSGNSLPEKNSEKNSLDHGQOSTLGNSVBCDKLDQSEI 600
 QY 1301 KSLMLCPYLIIKSMGSDALLFTYWNKASTSELMDFETISVCLHQFYWGKXYIA----- 1354
 Db 601 KSLMLCPYLIIKSMGSDALLFTYWNKASTSELMDFETISVCLHQFYWGKXYIANQEG 660
 QY 1355 -----RTGMMHARLQOLGSLDNLSTFNHSGHSDAVYHQSLEANI 1397
 Db 661 GTIVDRKSQTLPVSRNRTGMHARLQOLGSLDNLSTFNHSGHSDAVYHQSLEANI 720

QY 1398 TEVCLTALDTLSLFTLAFAKQGLADHGNPLMKVPEVYLCEFLQKHQSEALKNVETALR 1457
 DB 721 TEVCLTALDTLSLFTLAFAKQGLADHGNPLMKVPEVYLCEFLQKHQSEALKNVETALR 780
 QY 1458 SLIYKPFSTFYGRADMCAALCYEILKCNKSLSTRTASQULYFLMNNNDYGGKSF 1517
 DB 781 SLIYKPFSTFYGRADMCAALCYEILKCNKSLSTRTASQULYFLMNNNDYGGKSF 840
 QY 1518 VATHLQVITSVSQLADVVGIGTRFQOSLSIINNCAEERLIKHTSPSPSDVKDLTKRIR 1577
 DB 841 VATHLQVITSVSQLADVVGIGTRFQOSLSIINNCAEERLIKHTSPSPSDVKDLTKRIR 900
 QY 1578 TVLMAFAQMKHEHNDPEMLVDLYSLAKSYASTPELRLKTLWDSMARIHVNGDLSEAMC 1637
 DB 901 TVLMAFAQMKHEHNDPEMLVDLYSLAKSYASTPELRLKTLWDSMARIHVNGDLSEAMC 960
 QY 1638 YVHTVALVAEYLTRKNAVOMPEPLPHSHSACLRSRGVFPQCGTAFTVTPNIDEAS 1697
 DB 961 YVHTVALVAEYLTRKNAVOMPEPLPHSHSACLRSRGVFPQCGTAFTVTPNIDEAS 997
 QY 1698 NMEDVGMQDVHFNEDVLMELBECADGLMKAREVELIADYKLIIPYKRDPERLALH 1757
 DB 998 NMEDVGMQDVHFNEDVLMELBECADGLMKAREVELIADYKLIIPYKRDPERLALH 1051
 QY 1758 YDTLHRAVSKVTEVMSGRLLGTYFRVAFGQAQYQFTDESETDVEGFEDDGKEYIY 1817
 DB 1052 -----FEDDGKEYIY 1062
 QY 1818 KEPKLTPLSEISQRLKLYSDYKFGSENVAMIDSGKNPKDIDSKAYIQTHTVIFPDE 1877
 DB 1063 KEPKLTPLSEISQRLKLYSDYKFGSENVAMIDSGKNPKDIDSKAYIQTHTVIFPDE 1122
 QY 1878 KEOERKTEFERSHNRIRPFMPFQNGSKROGVEOCCRRITLAIHCPFYVKRIPY 1937
 DB 1123 KEOERKTEFERSHNRIRPFMPFQNGSKROGVEOCCRRITLAIHCPFYVKRIPY 1182
 QY 1938 MYQHTDINPIEVAIDEMSKVAELRQLCSSAEVDMIKLQLKQSSVSQVNAAGPLAYAR 1997
 DB 1183 MYQHTDINPIEVAIDEMSKVAELRQLCSSAEVDMIKLQLKQSSVSQVNAAGPLAYAR 1242
 QY 1998 AFLDITNTKRYPDNKVKYLKEYFRQVVEACGQALAVNEFLIKEDQLEYOEKMANREMA 2057
 DB 1243 AFLDITNTKRYPDNKVKYLKEYFRQVVEACGQALAVNEFLIKEDQLEYOEKMANREMA 1302
 QY 2058 KLSHIMEQICPLEKTSVLPNSLHIFNAISGTPSTNVHGMTSSSVY 2107
 DB 1303 KLSHIMEQICPLEKTSVLPNSLHIFNAISGTPSTNVHGMTSSSVY 1352

RESULT 14
 ABG61679
 ID ABG61679 standard; Protein; 1353 AA.
 AC ABG61679;
 DT 13-AUG-2002 (first entry)
 DE Cadherin-like asymmetry protein (CLASP) isoform #8.
 KM Human; autoimmune disease; haematopoietic disorder; Digorge syndrome;
 KM blood protein disorder; agammaglobulinemia; dysagammaglobulinemia;
 KM ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
 KM thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
 KM haemolytic anaemia; multiple sclerosis; rheumatoid arthritis; lupus;
 KM endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
 KM autoimmune pulmonary inflammation; organ rejection; inflammation;
 KM CLASP.
 XX Homo sapiens.
 OS
 XX
 PN WO200231117-A2.
 XX

PD 18-APR-2002.
 XX
 PF 15-OCT-2001; 2001WO-US32202.
 XX
 PR 13-OCT-2000; 2000US-0687837.
 XX
 PA (ARBO-) ARBOR VITA CORP.
 XX (GARM/) GARMAN J D.
 PA (CAND/) CANDIA A F.
 XX
 PI Lu PS;
 XX
 DR WPI; 2002-416861/44.
 XX
 PT New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
 PT an immune response, and for treating multiple sclerosis, rheumatoid
 PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
 PT and sepsis
 XX
 PS Disclosure; Figure 5A; 245pp; English.
 XX
 CC The invention relates to an isolated polypeptide (I) comprising an amino
 CC acid sequence that has 90 % sequence identity to one of the human
 CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E)
 CC sequences (PS). (I) is useful for identifying a compound or agent that
 CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
 CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for
 CC inhibiting a immune response in a subject. A pharmaceutical composition
 CC comprising a nucleic acid encoding (I), or (II) is useful for preventing
 CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
 CC the autoimmune disease is caused or exacerbated by increased activity
 CC of TH1 cells. CLASP-2 polynucleotides are useful as probes or primers for
 CC detection or inhibition of CLASP-2 expression (e.g., antisense or
 CC ribozyme-mediated inhibition), for gene knockout, etc. The CLASP-2
 CC polynucleotides can express CLASP-2 polypeptides, produce anti-CLASP-
 CC antibodies or are used as therapeutic polypeptides. The CLASP-2
 CC polynucleotide or fragments can be used in diagnostic (e.g., as probes
 CC for CLASP-2 expression), as a lymphocyte marker and for therapeutic
 CC purposes. CLASP-2 polynucleotides can construct transgenic and knockout
 CC animals, e.g., for screening of CLASP-2 agonists and antagonists. CLASP-2
 CC polynucleotides can screen for CLASP-2 agonists and antagonists.
 CC CLASP-2 polypeptides or polynucleotides can treat deficiencies or
 CC disorders of the immune system, by activating or inhibiting the
 CC activation, differentiation of immune cells and can treat or detect
 CC deficiencies or disorders of haematopoietic cells. CLASP-2 polypeptides
 CC or polynucleotides can increase differentiation and proliferation of
 CC haematopoietic cells, including the pluripotent stem cells to treat those
 CC disorders associated with a decrease in certain (or many) types of
 CC haematopoietic cells e.g., immunologic deficiency syndromes including
 CC blood protein disorders (e.g., agammaglobulinemia,
 CC dysagammaglobulinemia, ataxia telangiectasia, common variable
 CC immunodeficiency, Digorge syndrome, lymphopenia, thrombocytopenia, or
 CC haemoglobinuria). CLASP-2 polynucleotides or polypeptides can treat or
 CC detect autoimmune diseases, e.g., Addison's disease, haemolytic anaemia,
 CC Grave's disease, multiple sclerosis, rheumatoid arthritis, lupus,
 CC endometriosis, autoimmune thyroiditis, and autoimmune pulmonary
 CC inflammation. CLASP-2 can be used to treat anaphylaxis or
 CC hypersensitivity to an antigenic molecules, organ rejection or graft-
 CC versus-host disease (GVHD) and inflammation. ABG61670-ABG61708
 CC represent cadherin-like asymmetry protein (CLASP) sequences of the
 CC invention.
 CC
 CC Sequence 1353 AA;
 CC
 CC Query Match 62.0%; Score 6782.5; DB 23; Length 1353;
 CC Best Local Similarity 92.6%; Pred. No. 0;
 CC Matches 1324; Conservative 2; Mismatches 3; Indels 101; Gaps 3;
 CC
 QY 701 VLHHNPFYDEIKTELPTQLHEKHLLTFPHVSCNSKSGTKKRDVYETQVGYMWL 760
 DB 1 VLHHNPFYDEIKTELPTQLHEKHLLTFPHVSCNSKSGTKKRDVYETQVGYMWL 60
 QY 761 PLKXGRVVTSEOHIPVANSPLSGYGLGMGRHVGPEIKWVDGKPLIKISTHLVST 820

```

Db      61 PLKDKRVTTSSEHIVSANTLPSGYLGDELGWRGHYGPDKVDGKPLKISTLAVST 120
Qy      821 VYTDQHLNFPYCOCKTESSGAOLGNELVKYLKSHAMEGHVMTAFLTTIINQFRLVT 880
Db      121 VYTDQHLNFPYCOCKTESSGAOLGNELVKYLKSHAMEGHVMTAFLTTIINQFRLVT 180
Qy      881 RATOEEVAVNVTIRVLIHVVAQCHEEGLESLSYVYKAYKAEPPYVASEKYVHELTGSM 940
Db      181 RATOEEVAVNVTIRVLIHVVAQCHEEGLESLSYVYKAYKAEPPYVASEKYVHELTGSM 240
Qy      941 TTILKPSADPLTSNKLKTSWFFFDVLTYSMAOHLIENSKVKLLNORFPASTYHAAVEV 1000
Db      241 TTILKPSADPLTSNKLKTSWFFFDVLTYSMAOHLIENSKVKLLNORFPASTYHAAVEV 300
Qy      1001 VNMMLPHITQKFRDNEASKNANHSLAVFKRCFTFNDGPFVKQINNTYISCPAPDPT 1060
Db      301 VNMMLPHITQKFRDNEASKNANHSLAVFKRCFTFNDGPFVKQINNTYISCPAPDPT 360
Qy      1061 LEYKKEFLRWVCNHEHYIPLNLPMPFGKRIQRYODLDLDYSLTDEFCRNHFLVGLLR 1120
Db      361 LEYKKEFLRWVCNHEHYIPLNLPMPFGKRIQRYODLDLDYSLTDEFCRNHFLVGLLR 420
Qy      1121 EYGTALQERREVRLLAISVLKNLLIKHSFDDRYASSHQARIATLYLPLFGLLIENVQRI 1180
Db      421 EYGTALQERREVRLLAISVLKNLLIKHSFDDRYASSHQARIATLYLPLFGLLIENVQRI 480
Qy      1181 NRDVSPFPVNGMTYKDESLLPAPNPLVTPOKSTLNSLHKDLGASIGASPYTTS 1240
Db      481 NRDVSPFPVNGMTYKDESLLPAPNPLVTPOKSTLNSLHKDLGASIGASPYTTS 540
Qy      1241 TENINSVRNADSGSLISTDSGNSLPERNSEKNSLDKIQOOSTLGNSVVRCDKLQDSBI 1300
Db      541 TENINSVRNADSGSLISTDSGNSLPERNSEKNSLDKIQOOSTLGNSVVRCDKLQDSBI 600
Qy      1301 KSLMCFYLLKSMSDALFTYWNKASTSELMDFITSEVCLHOFQYMKRKYIARNOEGL 1354
Db      601 KSLMCFYLLKSMSDALFTYWNKASTSELMDFITSEVCLHOFQYMKRKYIARNOEGL 660
Qy      1355 -----RTGMMHARLQOLGSLDNLSTFNHSGHSDADVLHOSILEANITA 1397
Db      661 GRLVHDRKQOTLPVSNRRTGMMHARLQOLGSLDNLSTFNHSGHSDADVLHOSILEANITA 720
Qy      1398 TEVCLTALDLSLFTLAFKQNLADHGNPLMKVYEDVYLCELOKQOSETALKVFTALR 1457
Db      721 TEVCLTALDLSLFTLAFKQNLADHGNPLMKVYEDVYLCELOKQOSETALKVFTALR 780
Qy      1458 SLIYKPESTFYEGRAMCALCYEILKCNKSLSIRTAASQLLYFLMKNPDYTGKSF 1517
Db      781 SLIYKPESTFYEGRAMCALCYEILKCNKSLSIRTAASQLLYFLMKNPDYTGKSF 840
Qy      1518 VRTHLQVILISVSQLADVVIGIGTRFOQSLIINNANSDRLIKHSPSSDYDLTKRIR 1577
Db      841 VRTHLQVILISVSQLADVVIGIGTRFOQSLIINNANSDRLIKHSPSSDYDLTKRIR 900
Qy      1578 TVLMATAQMKHEHNDPEMLVDLQYSLAKSYASTPELIRKTMDSMARIHVNGDLSBANAC 1637
Db      901 TVLMATAQMKHEHNDPEMLVDLQYSLAKSYASTPELIRKTMDSMARIHVNGDLSBANAC 960
Qy      1638 YVHVTALVAEYLTRKAVQMBEPLLPHSACLRSGGVFROGCTAFVITPNIDEAS 1697
Db      961 YVHVTALVAEYLTRK-----GVFRQGCCTAFVITPNIDEAS 997
Qy      1698 KMEEDVGMQVHFNEEDVLMELLEOCADGLMKAEYELIADIYKLIPIYKRPDEFERLAL 1757
Db      998 KMEEDVGMQVHFNEEDVLMELLEOCADGLMKAEYELIADIYKLIPIYKRPDEFERLAL 1051
Qy      1758 YVTLRAVSKVTEVMSGRLLGTYPVAFPGQAQYQFTDSETVDEGEFEDGCKEYIY 1817
Db      1052 -----FEDGCKEYIY 1062
Qy      1818 KEPKLTPLSEISORLLKLYSDKFGSENVKMIQDSGKNPKDLSKAYIQTHTVIPFDE 1877

```

```

Db      1063 KEPKLTPLSEISORLLKLYSDKFGSENVKMIQDSGKNPKDLSKAYIQTHTVIPFDE 1122
Qy      1878 KELOERKTEPERSHNIRRFEMFPTQTKGKQGVBEQCKRTITLAIHCFPYVKRIPV 1937
Db      1123 KELOERKTEPERSHNIRRFEMFPTQTKGKQGVBEQCKRTITLAIHCFPYVKRIPV 1182
Qy      1938 MYOHTTDNPIEVALIDENSKKVAELRQCSSAEVDMIKLOLQSGSVQVNAAGLAYAR 1997
Db      1183 MYOHTTDNPIEVALIDENSKKVAELRQCSSAEVDMIKLOLQSGSVQVNAAGLAYAR 1242
Qy      1998 AFLDITNTKRYPDNKVKLLKEVFQFVACQOALAVNRLIKEDOLEVOEEMKANVREMA 2057
Db      1243 AFLDITNTKRYPDNKVKLLKEVFQFVACQOALAVNRLIKEDOLEVOEEMKANVREMA 1302
Qy      2058 KELEIMEHQICPLEKTSVLPNSLIHFNALSGTPTSTMVAGMTSSSSSV 2107
Db      1303 KELEIMEHQICPLEKTSVLPNSLIHFNALSGTPTSTMVAGMTSSSSSV 1352

RESULT 15
ABG61687
ID ABG61687 standard; Protein; 1353 AA.
XX
AC ABG61687;
XX
DT 13-AUG-2002 (first entry)
XX
DE Cadherin-like asymmetry protein (CLASP) isoform #15.
XX
KW Human; autoimmune disease; haematopoietic disorder; Digorge syndrome;
KW blood protein disorder; agammaglobulinaemia; dysagammaglobulinaemia;
KW ataxia telangiectasia; common variable immunodeficiency; lymphopenia;
KW thrombocytopenia; haemoglobinuria; Addison's disease; Grave's disease;
KW haemolytic anemia; multiple sclerosis; rheumatoid arthritis; lupus;
KW endometriosis; autoimmune thyroiditis; anaphylaxis; hypersensitivity;
KW autoimmune pulmonary inflammation; organ rejection; inflammation;
KW CLASP.
XX
OS Homo sapiens.
XX
PN MO200231117-A2.
XX
PD 18-APR-2002.
XX
PF 15-OCT-2001; 2001WO-US32202.
XX
PR 13-OCT-2000; 2000US-0687837.
XX
PA (ARBO-) ARBOR VITA CORP.
XX
PA (GARM/) GARMAN J D.
XX
PA (CAND/) CANDIA A F.
XX
PI Lu PS;
XX
DR WPI; 2002-416861/44.
XX
PT New human cadherin-like asymmetry protein(s) (CLASP)-2 for modulating
PT an immune response, and for treating multiple sclerosis, rheumatoid
PT arthritis, endometriosis, lupus, autoimmune thyroiditis, septic shock,
PT and sepsis
XX
PS Example 4; Figure 6B; 245bp; English.
XX
CC The invention relates to an isolated polypeptide (I) comprising an amino
CC acid sequence that has 90 % sequence identity to one of the human
CC cadherin-like asymmetry protein(s) (CLASP)-2 (hCLASP-2A, 2B, 2C, 2E)
CC sequences (PS). (I) is useful for identifying a compound or agent that
CC binds CLASP-2 polypeptide. An antibody (II) to (I) is useful for
CC detecting a CLASP-2 polypeptide in a sample. (II) is useful for
CC inhibiting a immune response in a subject. A pharmaceutical composition
CC comprising a nucleic acid encoding (I), or (II) is useful for preventing
CC or treating a CLASP-2 mediated disease e.g. an autoimmune disease, where
CC the autoimmune disease is caused or exacerbated by increased activity

```


THIS PAGE BEGINS BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:11:27 ; Search time 23.6546 seconds
(without alignments)
2620.804 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936

Sequence: 1 MSQPLLPASAEKTRKTRAL.....ISGTPSTWVHGHTSSSVV 2107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
- 2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
- 3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
- 4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
- 5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
- 6: /cgn2_6/prodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	359.5	3.3	1865	1 US-08-588-985-2	Sequence 2, Appl
2	359.5	3.3	1865	1 US-08-971-988-2	Sequence 2, Appl
3	199	1.8	3248	1 US-08-353-700-1	Sequence 1, Appl
4	199	1.8	3248	5 PCT-US95-16216-1	Sequence 1, Appl
5	177	1.6	3056	1 US-08-508-836A-8	Sequence 8, Appl
6	177	1.6	3056	2 US-08-629-001A-3	Sequence 3, Appl
7	177	1.6	3056	4 US-08-642-274D-3	Sequence 3, Appl
8	177	1.6	3056	4 US-08-952-127-3	Sequence 3, Appl
9	177	1.6	3057	4 US-08-952-014C-3	Sequence 3, Appl
10	173.5	1.6	3057	4 US-09-360-416-3	Sequence 3, Appl
11	173	1.6	3056	4 US-09-360-416-2	Sequence 2, Appl
12	172	1.6	3056	2 US-08-874-266-2	Sequence 2, Appl
13	167	1.5	10182	4 US-09-134-001C-3159	Sequence 3159, Ap
14	163	1.5	2285	1 US-09-308-375-2	Sequence 2, Appl
15	162	1.5	2482	1 US-08-328-254-6	Sequence 6, Appl
16	159	1.5	4536	4 US-09-180-422B-27	Sequence 27, Appl
17	158.5	1.4	1388	4 US-09-572-191-2	Sequence 2, Appl
18	158.5	1.4	1388	4 US-09-723-262-2	Sequence 2, Appl
19	158.5	1.4	1388	4 US-09-723-219-2	Sequence 2, Appl
20	157.5	1.4	2987	2 US-08-970-269A-29	Sequence 29, Appl
21	157.5	1.4	2987	2 US-08-970-269A-29	Sequence 29, Appl
22	157.5	1.4	3959	2 US-08-970-269A-30	Sequence 30, Appl
23	157.5	1.4	3959	4 US-09-407-562-30	Sequence 30, Appl
24	156	1.4	2329	3 US-08-755-587-16	Sequence 16, Appl
25	155.5	1.4	1786	4 US-08-973-462-8	Sequence 8, Appl
26	153.5	1.4	3418	2 US-08-639-501-2	Sequence 2, Appl
27	153.5	1.4	3418	3 US-09-044-946-2	Sequence 2, Appl

28	153.5	1.4	3418	3 US-09-044-908-2	Sequence 2, Appl
29	152.5	1.4	1038	4 US-09-541-782-4	Sequence 4, Appl
30	152.5	1.4	1038	4 US-09-723-820-4	Sequence 4, Appl
31	152.5	1.4	3418	3 US-08-755-587-44	Sequence 44, Appl
32	151.5	1.4	3418	2 US-08-603-753D-4	Sequence 4, Appl
33	151.5	1.4	3418	4 US-09-099-753-4	Sequence 4, Appl
34	151.5	1.4	3418	4 US-08-986-106-4	Sequence 4, Appl
35	148.5	1.4	1211	4 US-09-134-001C-4820	Sequence 4820, Ap
36	147	1.3	3144	2 US-08-457-273B-42	Sequence 42, Appl
37	147	1.3	3144	3 US-08-556-419-21	Sequence 21, Appl
38	147	1.3	3144	4 US-09-041-886-15	Sequence 15, Appl
39	146.5	1.3	3079	5 PCT-US94-00198-4	Sequence 4, Appl
40	146	1.3	2818	4 US-09-542-331-2	Sequence 2, Appl
41	146	1.3	2818	4 US-09-510-791-2	Sequence 2, Appl
42	145.5	1.3	1447	4 US-09-376-330-17	Sequence 17, Appl
43	144	1.3	3144	1 US-08-246-982A-6	Sequence 6, Appl
44	144	1.3	3144	1 US-08-453-265-6	Sequence 6, Appl
45	143.5	1.3	988	3 US-08-851-843A-69	Sequence 69, Appl

ALIGNMENTS

RESULT 1
US-08-588-985-2
; Sequence 2, Application US/08588985
; Patent No. 5777094
; GENERAL INFORMATION:
; APPLICANT: MICHIOYUKI MATSUDA et al.
; TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESSES:
; ADDRESSES: Wenderoth, Lind & Ponack
; STREET: 805 Fifteenth Street, N.W., #700
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,985
; FILING DATE: January 19, 1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-8850
; TELEFAX:
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1865 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL:
; ANTI-SENSE:
; FRAGMENT TYPE:
; ORIGINAL SOURCE: spleen cell of homo sapiens
; US-08-588-985-2
Query Match 3.3%; Score 359.5; DB 1; Length 1865;
Best Local Similarity 18.7%; Pred. No. 5.4e-21;

Matches 393; Conservative 345; Mismatches 764; Indels 605; Gaps 99;

QY 136 ROLPNKVKLDKPLVHYVEEDVEEDKEDDASISGQKGI-TGKHGTYKGNMMSAISVTM 194
 Db 3 RWPPTK--REKYGVAFYNDAR-GADELSLOIGDVIHILEYEGM-YRG-----YTL 51
 QY 195 RS-----FKRRFPHL---IQLDGSKYKFFELDKLOKRP-----KGSIFLFLYG 235
 Db 52 RKSKKGIKIPASYSIHLKEAIVEGKGQHEVIYIPGDL---PLIQEVTTTLRWSWTIMQOLYV 108
 QY 236 VSRFRN--NKVRRAFELKMDKSSYL---LAASEVEMEMETITL-----NKIQLFEEA 286
 Db 109 QDNREFRSVRHMITYDL-IEKRSQILSGTLPQDELKELKKVAKIDYGRILLDLDLVR 167
 QY 287 MOEKRGDSDHEDEQSLGSGSLDLYPELAK-----SARAEIKLSESRYK 336
 Db 168 -----DE-----DGNILD---PELSTISLPRAHIAKQYERLQEEKSQK 206
 QY 337 LFLYLDPAQKLDPS-----SABPEVKSFEERKFRILVKCNDSFNLQCCVAENEBGPTT 391
 Db 207 -----QNIIDINROAKFAPATPSLALF-----VNLK-----NVVCKIGDAB----- 241
 QY 392 NVPEFFYTLISLPIKXNRKISADPHVDLNFVSVRQMIATTSBALMNGSGEPTQSALRGIL 451
 Db 242 -----VLSLSDYDVESKFISENYLVKRWSSSGLPKDI-----DRLHNLRAVF 282
 QY 452 HEAMQYPKQGISVTCPHPDIFLVARIEKVLQGSITHCAPYKMSDSSKVAQVILKVA 511
 Db 283 TDLGSDLKREKISVPC---QIVRGME-----LRDNTRKLTLSGL----- 321
 QY 512 KQACORLGOYRMPAPMAAKTLFQDASGNLDKNAFSAIYRQ---DSNKLSDNDMLKLAD 568
 Db 322 -----RRPFGVAVMDVTDIINGKVDDEKQHFIPQPVAGENDFLQTYINVIYIA 371
 QY 569 FRKPEKMAKLVILGNLDITIDNVSDPEFYVNSVSIPTKQFETCSKPTTFEEVEEFVVC 628
 Db 372 KEVNHKGQGLWVTLKLPDIIHQIRKEFPHLV-----RTTAVARKTG----- 414
 QY 629 IPKHTQPTTYTNHLVYVPKYLYKSDQSEPAKARNIACIEFKDSEDESOPLKCIYGRP 688
 Db 415 FPEIIRPGDV-RNDIYVTLVQSDPD-KGSKTTAKNVEVTVSVVD---EOKKRLHENI-PP 468
 QY 689 GGEVFTRSAPAAVLHNN-QNPEFYDEIKELPQOLHEKHHLLTFPHVSCDSSKSGSTYK 747
 Db 469 GAGDEAISEKYSYIYQVKQPRWFETVYKAIPIEDVNRSHLRFTHRRSSQDS----- 521
 QY 748 RDVVEITOVQVSMPLK-DGRVVTSSQH-----IPVANIPSGVLGVOELG 792
 Db 522 KQSEKIFALAFVKLMRYDSTTLRDSHDLIYVKAELKLEDAATYLSLSTYAELEEK 581
 QY 793 MGRHVGPEIKWVDG---GKPLKISTHLVSTVYTDQHLNFPQYOCQKTESGAQALGNEL 849
 Db 582 HSA-TKSNQSLSCSITISKDSPOISTLVCTKLTQVNDLGLKMKMSNT-----SLL 632
 QY 850 VKTKLSLHAMEGHVMIATFLPTIINQJFRVLTRATQEE---VAVNTRVYIIVVAQCHEE 905
 Db 633 QOQRLQMLKVDGGEVVKFLQDLDLALFNIMMENSESETEPDLVFLDVLFIIGLIAIRKEFO 962
 QY 906 GLESHRSYVKAUYKA-----EPVY-ASEKYTHAEELTKMTIILKSNADPLFS 953
 Db 693 HRPVLETIYIKGHFSATLAVTKLTKVKNVDGAEKPRGVAEQYKMKMA-LESTIEFFIYR 751
 QY 954 NKLKTSWFFFDVLISMAOHLIENSKVLLRNQRPASYNHAAVETVVMMLMHIITQKR 1013
 Db 752 SRL-----FNQLYENKGEADFVESLLOLFRS-----INDMSSMSDQ-- 789
 QY 1014 DNEPASKANANSLAVETIKRCFTMDRGFVFKQ---INNYISCFADGDPKTYFEYKFE 1069
 Db 790 -----TVRVGAALIKYLPITVNDVKLVF---DEKELSKMFEETI 825
 QY 1070 RVCNHNHYIPLNLPMRFGKGRIGR-YODLOLDYS--LTDPECRNHFVGLLREVGTL 1126
 Db 826 -----LNVPM--GLTTIQKLYCLEIYHSDLFTQHDGR-BILLPMITDQKLYH 871

QY 1127 QEFREVRLLAISVLKNLLIKSIFDDRVASRSHQARIATLYPLFGILLNVOIRINRVDS 1186
 Db 872 ERQEDLEACC-----QLLSH-----IIEVLVR---KQVG 897
 QY 1187 PPPVNAGMTVDKESLALPAVNDLPTPQKSGTLDNSLHDLGALGASIPYTTSTPNINS 1246
 Db 898 P-----TORHVOIIM-----EKLJLTVA----- 915
 QY 1247 VKNADSRGLISTDSGNSLPERNSEKNSLDHGQOSTTSGNSVVAQKLDQSEIKSLMC 1306
 Db 916 -RTVISMG-----RSEILIGNFVA-----C 934
 QY 1307 PLYIILKSNDDALFTYWNKAST-----SELMDFPTISEVCLHOFQMGKRYIARTGMMA 1361
 Db 935 MPAIRKQED---YHAAHLIKTPGMRITDVDF-LMEFFIMFKNLIGKNVPRFDVVI-- 987
 QY 1362 RLQQLGSLDNLSTFR--HSYGHSDADVIHOSLL-EANITATEVCLTALDTLSLFTLAFKQ 1418
 Db 988 -----MMVQNKVFLRAINQY---ADMINKKFLDAQNFELQ----- 1020
 QY 1419 LLDHGHNPMLKCVVDVILCFQKHQSESTALKNVFTALRSIL-----YKFPST 1466
 Db 1021 -----LMNNYFHLAVAFIT--QESILOENFSSAKRAKILNKYGDMRROIGFELIDM 1069
 QY 1467 FY---EGRADMCALCYEILKCNKSLSIRTEASQULYFLMRNNFDYTGKKSFYRTHLQ 1523
 Db 1070 WYNLOGHKIKFIPENWVGPILEWTLIPETELRATGPIPFDMWGCFHSF--RSFGMFENE 1127
 QY 1524 VIISVSQILADVVGIGGTRFOOSL--SIINNCANSDLIK-HTFSSDYADULTRI--- 1576
 Db 1128 II---TKLDHEVEGGRGBOYKVLFDKILLEHCRGHKYLAKGETFVKLVARMLRMLDY 1184
 QY 1577 RTVLMAQVOMKEHNDPEMLVDLOYSLAKSY--ASTPELRKTMBSMARIHKONDLSEA 1634
 Db 1188 RTIM-----HDKENKRNSTYVNVANFYKEIREEMYRIYKLCDLHKECNYTEA 1236
 QY 1635 AMCVUHVTAIVAEYLTRKEAVQMEBPLLPHSHACLRR--SRGVPFROGCTAFRVTENI 1692
 Db 1237 AV-----TLLHAKLLKM-----SEDCVVAHLTQDQ-----YQVTT--- 1268
 QY 1693 DEEASIMEDVGMQDVH-FNEDVLMELLEQCADGLMKARIELADIYKLIIPYKRD 1751
 Db 1269 --QOGLKQLQVOEIIHYFDKGMME--BAIALGKELAOYE-----NEMFDY 1311
 QY 1752 BRALHUYTLHAYSKTEVWMSGRRLGTFRVAFPGQAQYQPTDSETDVEGPFEBD 1811
 Db 1312 EQLSELKKQAOFYENIKVI---RKPEDIFAVGYIQ-----GPTPLR 1353
 QY 1812 GKEYIYKEPKLTPLEISQRLKLYSDKFGSENVMIQDSGVNPKDL-DSKAYAIQVTH 1870
 Db 1354 GKVFYIRGKEYEPRDFEARLLTQFPN--ABKMTTSBPG---DDIKNSPGQYIQCF 1406
 QY 1871 VLPFPD-----EKLEQERKTEPESHNIRRMFEHPTQOTGR--QGVVEBOCKRTIILTA 1924
 Db 1407 VPKPDLDPKPRPVPSEQIVSFYRVNEQREYSPRIKRGKGNPNDEANMMIETKIYTT 1466
 QY 1925 IHCPPYVKKRIPVWQHHTDNPJEAIDEM---SKVAELRQLCSSAEVDMIKLOKL 1980
 Db 1467 AKLDPGILRWPEVSVFVWEISPLENALETQLTNDKINSVQOHLDDPSLPINLSML 1526
 QY 1981 QGSVSVQVNAAGLAVARAFLDYNTKRYPD--NKVKLKEVFRQFVEACQALAVNERLI 2038
 Db 1527 NGIYDPAWGGFANYEKAFTDRYLOEHRBAHEKIKKQDLIAMIPLAGSIRIHGXV 1586
 QY 2039 KEDQLETOEBMKANTRENAKELSEIMHQCPLSEKTSVLPNSLHIFNAIGSTPISTVNH 2098
 Db 1587 TEALRPFHRMRBACFKQKEKEK-----EYGVRIIMSSSL--DDRGSRSRSNVR 1634
 QY 2099 GWTSSSS 2105
 Db 1635 SFTMPSS 1641

RESULT 2
US-08-971-988-2
Sequence 2, Application US/08971988
Patent No. 5786461
GENERAL INFORMATION:
APPLICANT: MICHIOKI MATSUDA et al.
TITLE OF INVENTION: CDNA OF DOCK180 GENE AND DOCK180 PROTEIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/971,988
FILING DATE: 17-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/588,985
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-8850
TELEFAX:
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1865 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: protein
HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE: spleen cell of homo sapiens
US-08-971-988-2
Query Match 3.3%; Score 359.5; DB 1; Length 1865;
Best Local Similarity 18.7%; Pred. No. 5.4e-21;
Matches 393; Conservative 345; Mismatches 764; Indels 605; Gaps 99;
QY 136 ROLPNKVVKLKLPVHVVEDEVDKEDDASISGQKGI-TKQGLYKGNMNSAISVTM 194
DB 3 RWPPTK-REBKGVAFVYVDAR-GADELSLQIGDTVHILETYEGM-YRG-----YTL 51
QY 195 RS-----FKRRFHL---IQGDGSKYKFEFLKDLQKPE-----KGSIFLGLYNG 235
DB 52 RKSKKGIIPASTIHKEAIVEGKGHEVPIPGDL---PLIOEVTTLREWSITWMLYV 108
QY 236 VSEFRN--NKYRRFAPFLKQDKSSYL---LAADSEVMESEWITIL-----NKILQLNFEAA 286
DB 109 QGNREFRSVRHMIYDL-IEWRSQILSGTLPODELKELKRYAKIDYGNRIIDLDLYVR 167
QY 287 MGEKNGSDHEDDEGSLGSGGLDYSYBELAK-----SAREAEIKIKSSRYK 336
DB 168 -----DE-----DGNILD---PELTSTISLFRHAIARAKQVEERLERLEKESQK 206
QY 337 LFLYLPDAQKLPDS-----SABPEVKSFEKPKGRILVVCNDSFYLQCCVANESEGPTT 391
DB 207 -----QNIDINROAKFPAATPSIALF-----VNLK-----NVVCKIGDAE---- 241

QY 392 NVEPFVTLSTLEFDIKYNRKISADPHVDLNHRSVRQMIATTSBALMGSGPETQALRGIL 451
DB -----VLMSTLYDPVESKFISENYLVWSSSGLPRDI-----DELNHLRAVIF 282
QY 452 HEAMQYKQGIFFSTGCHPDIFLVARIKVLQSGITHCABPYMKSSSSKVAQKVLKRA 511
DB 283 TDLGSKDKIREKISFVC---QIVRVGRME-----LADNVTIKLTSGI----- 321
QY 512 KQACORLQGYRMPFMAARTLFEKASGNLDKXARSAIYRQ---DSNLSNDMDMLKLAD 568
DB 322 -----RRPFVAVAMDYTDIINGKYDEDEKHFIPFQPVAGENPFLQIVINKVIAA 371
QY 569 FKPEPKAKLPVILGNLDITIDNVSSDPENYVNSSYIPTKQFETSKPTIFVEVEPVPC 628
DB 372 KEVNHKGGLWTLTKLPDGHQIKKEPFLVD-----RTAVARKTG----- 414
QY 629 IPKHTQPTTYTNHLYVPKYLKVDQSGSPAKARIAICIEFKDSDEDSQPLKITYGRP 688
DB 415 FPEIIMPDDV-RNDIYVTLVQGDPD-KGSKYAKKVEVTSVYD---EDGKRLRHVI-FP 468
QY 669 GGPVFTSAPFAVLHHH--ONPEFYDEIKELPTOLHEKHLLLPFHVSCONSGSGRYK 747
DB 469 GAGDEAISEKSVITYYQKQPMFETVVAALPIEDVNRSHLRTFRHSSQDS----- 521
QY 748 RDVETQVGSWLPPLK-DGRVYTSQR-----IPVSAMLPSCYLGYOBIG 792
DB 522 KQKSEKIFALAFVVKLMRYDGTTLRGEBHDLVYKKAACKLEDAATYLSLPSTKALBEKG 581
QY 793 MGRHYGPEIKAVDG---GKPLKISTHLVSVYTYDQHLNFPQYCKNTESGAOLGHEL 849
DB 582 HSA-TGKSMQSLGSCCTISKDSFQISTLVCSCTLTONVLLGLLKRSTNT-----SLI 632
QY 850 VVYLSLHAMEGHWVIAFLPTINQFRVLPRATOE---VAVNVTVIIVHVAQCHEE 905
DB 633 QGNLRQLKVDGGEVVKFLQDTLDLALFNIMMENSEFPTDLVFPALVPIIGLIDRKFQ 692
QY 906 GLESHLRSYVVKYAKA-----BPVY-ASEYKTVHEELTSMYTLIKPSADFLTS 953
DB 693 HENPVLETYIKHFSATLAATYKTLKLVKNYDGAKEPQVNBGLYAKMA-LESIFKPIVR 751
QY 954 NKLKYSWFFPDVLIKSMAOHLIENSKYKLLBNQRFPAHYHNAVETVNMIMPHITOKFR 1013
DB 752 SRIL-----FQVLYENGEADFVESLQLPFRS-----INDMSSMSDQ-- 789
QY 1014 DNEPASKNANSLAFIRGCTFMDRGFVFKQ---INNYISCFARPDKTLFEKFEEL 1069
DB 790 -----TVRVGALKLXVLPITVNDVXLPF---DPKELSKNFTPEI 825
QY 1070 RVVCHNHHYIPLNLPMPFGKRIQR-YODLDLDS--LTDFECRNHPLVGLLREVGTA 1126
DB 826 -----LNVFM--GLLTQKLYCLIEIYHSDLFTQHDR-ELLPMPTDQDKYHL 871
QY 1127 GEFREVRLLAISVLKNLLIKISFDDRVSASHQARIATVYLPFGLILENQRINRVDVS 1186
DB 872 ERQEDLEACC---QLLSH-----ILEVLYR---KQVG 897
QY 1187 PPPVNAQMTVDESALAPVAVPIYTPQKSTLDSNHLKDLGASIGIASPYTTSFNPINS 1246
DB 898 P-----TORHVQIIM-----EKLRTVA----- 915
QY 1247 VNAADRSGLISTDSGNLPERNSEKNSLDKHOQSSTFGNSVVRCDKLQSEIKSLMLC 1306
DB 916 RTVISMG-----RSESLIGNFVA-----C 934
QY 1307 FLVYIKSMSDALFTYWNKAST---SELMDFTTISVCLHOFQYMGRIYARTGMMA 1361
DB 935 MTAIRQWED--YHAAHLIKTFGMRRTVDVDF-LMETPIMFKNLICKNYVPRPDWVI-- 987
QY 1362 RLQQLGSLDNLGTFN--HSYGHSDADYVHQSIL-EANITATEVCULADLTLSIFLAPKQ 1418
DB 988 -----NNMYQNKVFLRAIQY---ADMINKKFLDADNELQ----- 1020
QY 1419 LLAHDGHNPLMKKVPDVVLCFLQKHQSEETALKVNFALRSIL-----YKFPST 1466

```

Db 1021 -----LMNNYFHLAVAFLT--QSLQLENFSSAKRAKILNKYGMKROJGFELRDM 1069
Qy 1467 FY---EGRAMCALCEILKCCNSKLSIRTEASOLYFLMNNPPTGKSPVATHQ 1523
Db 1070 WYMLGQHKKIFIPBMBGPILEMTLIPETELRKATIPFFPMOCCEFFST--RSFQWENE 1127
Qy 1524 VIISVSQIADVAGIGSTRFQOSY--SIINCCANSRLIK-HTSFSSVDYKDTTKRI-- 1576
Db 1128 IT---TLDHHEVSGRDEQYKLPFDKILLEHKKHLYLAKTGETFYKLVVRLMELLDY 1184
Qy 1577 RTVLMATAQMKHEHDEBMLVDLOYSLAKSY--ASTPELKTWDSMARHVNQDLSXA 1634
Db 1185 RTIM-----HDENKEMNSCTVYANLYNFKIEIREEMRYLYKLCIDHKECDMYTEA 1236
Qy 1635 AMCVHTALVAEYLTRKEAVOWEPRLPHSHSACLR--SRGVPFQOGCTARVITPNI 1692
Db 1237 AY-----TLLHAKLLKM-----SEDCVAHLTORJG-----YQATF-- 1268
Qy 1693 DEBASMEVDYGMODVH-FNEVDLMELLEOCADGLMKAREYELADIYKLIIPYERKRD 1751
Db 1269 --QGOLKEQLYQELIHFDKGMWE--EALALGKELAEQYE-----NEMFDY 1311
Qy 1752 ERLAHLVDTLHRAYSKVTYVMSGRRLGTFRVAFEGQAQYQFTDSETVGEPFEDD 1811
Db 1312 EQLSELKKQAOQFENIVKVI---RPKPDYFAVGYGQ-----GPPFLR 1353
Qy 1812 GKRYIKERKLTPLSEISQBLKLYSDKPSSENVKMIQDSGXNPKDL-DSKAYIQVTH 1870
Db 1354 GKPFITRGKEYEPREDEARLLTOPRN--AEKMKTTSPG--DDIKNSPGQYIQCTF 1406
Qy 1871 VDPFD-----EKELOERKTEPERSHNIRREMEFPTQGR--QGVBEQCKRRTILTA 1924
Db 1407 VKRKLDPKPHNPVSQIVSPFRVNEVQGFESRPIRKGCKNPDMEFAMWERTITYT 1466
Qy 1925 IHCFPVYKRIPYMYQHTDNLNPLEVAIDEM---SKCAELQOLCSSAEVDMIKIQLK 1980
Db 1467 AYKLPGLRFEVYSVFMVVISPLENAIETMQLTNDKINSMVOQHLDPSLPINPLSM 1526
Qy 1981 QGSVSQVYVNGPLAVARAFLDNDNTRKYPD--NKVKLKEVPFQVAFBAGQALAVNERLI 2038
Db 1527 NGVDPVAVMGFPANVECAFPTDYLQEHPEAHEKIEKLDOLIMQIPFLAEGIRINGDV 1586
Qy 2039 KEDOLEYOEEMKANYEMAKELSEIMHEQICPLEEKTSVLPNSLIHFNASTGPTSTWY 2098
Db 1587 TBLRPHHEMBAQCFQKKEKVEK-----EYGVRLMPSL--DDRGRSRRPSMVR 1634
Qy 2099 GMTSSSS 2105
Db 1635 SFTMPSS 1641

```

```

RESULT 3
US-08-353-700-1
Sequence 1, Application US/08353700
Patent No. 5599919
GENERAL INFORMATION:
APPLICANT: YEN, TIMOTHY J.
APPLICANT: RATTNER, JEROME B.
TITLE OF INVENTION: NUCLEIC ACID ENCODING A
TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN.
TITLE OF INVENTION: AND METHODS OF USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
STREET: 1601 MARKET STREET, SUITE 720
CITY: PHILADELPHIA
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

```

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,700
FILING DATE: 09-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: REED, JANET E.
REGISTRATION NUMBER: 36,252
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN

```

```

Query Match 1.8%; Score 189; DB 1; Length 3248;
Best Local Similarity 18.5%; Pred. No. 1e-06;
Matches 424; Conservative 342; Mismatches 828; Indels 700; Gaps 107;

```

```

Qy 94 CSTVPAKAEENAS-----LFTVCITNTYSDMHLVNYKYEDYSGFEROLPN 140
Db 121 CSELEERSQQAASADVSLNCPCTPQKIFTPPLTPSQ-----YSSGSKYEDLEKY-----N 172
Qy 141 KYVXLDKLPVHYEYDEEDVDKDDASLSGQCG-----IRKH-----GWL 181
Db 173 K-----EVEERKLEAEVYKALQAKASQTLPOATNHRDIAHQASSVFSNQ 220
Qy 182 YGNMNSAIVTMRSEFKRFFHLIQDGSYFBEKDLQKEPK-----GSITFL 230
Db 221 QKTSHTLSSNSQRPPIRDF-----SASY-----FSGLEVTPTSSTIQIGRDNASSPF 271
Qy 231 G-----FLYVSPNNKVRPAFELKQODSSYLLAADSEVEMEWITLTKI--LQLN 282
Db 272 GNSSPHLLDQKAQNOELRNKINLELR-----LQGEHEKMGQ-----VKKFQELDQ 321
Qy 283 PEAAMQKRGNSHDDQSKLESGSGLDVLPBLASAR---EALIKLS--EERYLKF 338
Db 322 LEKAYEL-----IEKEV-----LHKCRDELVRTAQQDASTKYTALQKXKL 367
Qy 339 YLDDPAQKLDFFSA-----EPVYKSPFEKGRILVKNCDLAFNLQCVAVENEBGPTTVE 394
Db 368 TEDLSQQRNAMESACSLQKIKKEKEKFEDEL-----SROQ 404
Qy 395 PFFVTLSLFDIKYNNKISADPHVDNHSVQ---MIATTSPALMGSGPETGALRGI 450
Db 405 RSFQTLDECIQMKARLQELQAKNMENIVQAEIDKLTYSVQOLENN----- 452
Qy 451 LHEAMQVFKQGISVTCRPHDILVARI--EYVLOGSITHC--AEPYMSSDSQKAVQV- 507
Db 453 -----LEBPQKL-----CRABQAFQASQIKENELRSMEEKENLLKSHSEQAFREVC 503
Qy 508 -----LKNAKOA-----CORLGYRMPFAWAARTLPDQASGNDLKNARFSAIYRQDNKLSN 559
Db 504 HLEALKIKIQCLNQSQPAEMKAKNTSOETMLRDLQEKIN-----QDNSLTL 553
Qy 560 DDMKLADLDFRK-----PEKNAKLPIVILGNDDITIDVNSDPFNNYSSSYPTQFETCS 614
Db 554 EKLKLAVALDLEKQDCQDLKKREHNIQGLNDKLSKTEKESKALLSALBKKEYEELK 613
Qy 615 KTIPTF-----EVEFVPCIRKHTQPTTYITNHLVYRK-----YIKYSQ 655
Db 614 EKKTLFCWKSXENKLTLMQSEKENLQSKINHETCLKTQOIKSHBYNRYRTLEMDRE 673

```

656 KSPAKANAIACIEFDSDEBDSOPKCIYGRPGGVFTTRSAFAVLLHHNHNEFYD--- 712
 674 NLSVEIHNHNHNLDSK-SVEVEFOKL-----AYMELOQKAEFFDQGH 714
 713 --EIK-----IELPQHEKHLLLTFRHVSNDNSKSGSKKRDVVEQVGSMLPLL- 763
 715 QKEIEMCKTQSQTQOVDELEKLOL-----SNEIMEKDRCYQDILAEESLRDL 767
 764 --KQGVNTSEOHIPVSAHLPBGYLGYOELGMORHNGPEIKWVDGKPLIKTSTHVVSTV 821
 768 KSKDMSLVNREDB-----QKSLNLPDOOPAMHNSFANITIGQGSMPSEBSECLBA-- 818
 822 YTDQOH-----LHNEFYQCKTSGAQLAGNELVYKLALHMEGHVNAIPLTLLNOL 875
 819 ---DQSPKNSAIIQNNVDSLEFSELSQKQNSDLQCESELVQIKGEIENIMKA--EQM 873
 876 FVLVLRATQBEAVANVTYIHH--VVAOCHEBGLBHLASVYKAVKAPYVASEKTYH 933
 874 HOSPAVITSQRIKQEDTSAHQNVAA-----TJSALENTEKELOLLMDKYTEQAEI- 927
 934 BELTSMNTTILKPSADFLTSNKLKYSWFFPVLKISMAQHILNISKVLLNORPASY 993
 928 QELK-----SNHLEBLSKELQLLSTLSLEKKENSITSLKRE----- 968
 994 HNAVETVNMMLPHITQKFRDNEBASKNANSHLAVEIKACITFMDRGFVKOINNYISCF 1053
 969 -----IEELTOENGTLKEINASINQKQNLQKSESPA-----NYIDBR 1007
 1054 ABGDRTLEFYKPEPLRVACNENHYPILNMPFGKGRIOF-----YODLODYSLTD 1106
 1008 EKSISSLSDOYKQEKIL-----LOKCEETGNAVEDLSQKXKAAQ 1047
 1107 EFCRNHFLVGLLREVTALQEFREVRLLAISVLKYL-----IKHSPDRVASHQARIAT 1164
 1048 E---KNSKLECLNLC-TSLCENR-----KNELEQLKEAF-----AKHQEFLTK 1088
 1165 LYLPLFG-----LLIENVORINRVDSPPPVN---AG-----MTVKDESIALPRAV 1206
 1089 L--AFAEBRNQNLMELETVQOALRSEMTDQNNQNSKSEAGLQKQIMTLKEE----- 1138
 1207 NBLVTPQKSGTLDNSLHK--DLGALISGLASPYTT-----STPNINSRNADSRGS 1255
 1139 -----QNKQKQKENVNDLLOENBOLMKVMKTKHBCQNLSESPINNSVAKERESEBN 1186
 1256 L-----ISTDSGNS-----LPERNSKSNLSDHQOOSTLGN-S 1288
 1187 QCNFRQMLBVEKESLSDSYNAQULVLEMLRNKELKLOESBEKEXCLOHELOTRGDLE 1246
 1289 VVRCDKLDQSEIKSLMLCPL---YI-----LKSMSDPAFTYWNKASTSELMDFPTI 1337
 1247 TSNLQOQSGOEISGLKDCIDAEKYSISGPHLSSTQNNNHLQCSLOTTNNKLANLEKI 1306
 1338 SEVCLHOFQYMGKRYIARTGMNAR-----LOQJGSLDNLSTFNHNSYGHSDADVL 1387
 1307 CEI-----LOAEKELVTELDNRSECTATRTKMAEVEGCLNBEVKIL-----NDOSGLL 1356
 1388 HOSLLE-----ANATEVCALALD-----TJS-----LETLAF 1415
 1357 HGBELVDIDIGGERGEOPNEOHVPSLAPLBESNSYEHLTISJKEVOMHMFLOEKLISLOS 1416
 1416 KQQLADHGNHPLMKKVPDVYLCELOKHOSETALKNVF--TALRS---LIYKFPSTYE 1469
 1417 EKHILHDO-HCQMSKMSSE---LOTYVDSLKAENLVSTJNLRNROGLVEMOUGLE 1470
 1470 GRADMCALCYELKCNKSLGSIIRTEASQOLYTLMRNPFYTGKGSFVTRTHLOVITISVS 1529
 1471 GLVPSLSSSCVP---DSSSLSSLSGDS-----FYRALLQOTGDSL-SNIEGAVSAN 1519
 1530 QLIADVVGIGTRFOQSLIINNCAUSDRLIKHTSSSDVKDITKIRVLMATAQMKH 1589
 1520 QCSVDEV-----FCSSLQTYVDSLKAENLVSTJNLRNFOJDLVK----- 1558
 1590 ENDEPMLVDLOYSLAKSVASTPELRKTWLDMA-----RIHVKNGDLS----- 1632

1559 -----EMOJLEBGLVPSSLSSCVPPDSSLSISGDSFYRALLQOTGDSLNSNLEGVSA 1614
 1633 --EAMCVYHVAIVAEVLTTRKEAVQWGPPLPHSHASACLRSGRGVFPQCGAFVITPN 1691
 1615 NQCSVDEVPCCSLQEBENLTRKETPSAPAKGVEELSCL-----EVRQO-----S 1658
 1692 IDEBASMEVDVQMOVHFNEBDMELMELQCADGLMKARVEL--IADYKLIIPYEKRR 1749
 1659 LEKEBKESQOIMK--NKEI--QELQO-----LSSRQELDCIKRY-----LSEMQ 1704
 1750 DFERLAH--YDILHRAYSKTEVMA-----SGRLLIGTYFRAVAFQQA 1791
 1705 WQOKLTSTVLEMSFLAEKQTEQLSLEVARLOGLDLSRSLG----- 1753
 1792 AQOYFTSEIDVEGFEEDE--CKEYIYKPELTPLSLSQRLKYSKDPSE-----NVK 1846
 1754 ----IDTEDALQGNHSCDISKEHTSETTEPTPHVDHQ-----ICKDAQODLNLDE 1803
 1847 MIDSGKYNP-----KDLDSKYAVIQVTHVIPPDEKELOERKTEPERSHNRPMPE 1899
 1804 KITETGAVKPTGECGSGSPDNTYR-----PGEDEKQSSSICISLSTSGNALVP 1855
 1900 MFTQGTGRQCGVEBQCKRRITLAIHCFPPYKRIPVMYQHTDNLNPIVAIDEMSKV 1959
 1856 MDFL-----GNQED-----IH-----NLQLRVKTSENENLRLHV-IEDRDKV 1893
 1960 A-----ELAQOLGSAVDMIKLOLQKQGSVQVANGPLAVARAFDIDNTKRYPMNKYL 2015
 1894 BSLNEMKELDSKHLQEVQMLTKTEACTIEKTVIGELKENSJLSE-KLEVPSCDDEL 1952
 2016 LKEVFRQFVEACQALAVN-ERLIKED---OLEYQEBKANYREMAELSEIMEQICP 2070
 1953 LQRV--ETSSEGLNLSLMEHADKSSREDIGDNVAKNDSEKRFIDVENHLSIR----- 2004
 2071 LEEKTSVLNLSLHI 2084
 2005 -SEKASIBHEALYL 2017

RESULT 4
 PCT-US95-16216-1
 ; Sequence 1, Application PC/TUS9516216
 ; GENERAL INFORMATION:
 ; APPLICANT: Yen, Timothy J.
 ; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
 ; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
 ; NUMBER OF SEQUENCES: 4
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dann, Dorfman, Herrell and Skillman
 ; STREET: 1601 Market Street Suite 720
 ; CITY: Philadelphia
 ; STATE: PA
 ; COUNTRY: USA
 ; ZIP: 19103-2307
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US95/16216
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/353,700
 ; FILING DATE: 09-DEC-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Reed, Janet E.
 ; REGISTRATION NUMBER: 36,252
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (215) 563-4100

Db 1705 WOOLITSTLEMESKLAERKOTBOLSELEVARLOGLULSSRLG----- 1753
 Qy 1792 AQOYFDTSEDFEPEDED-SKEYIYKEPKLTPLEISORLLKLYSDKFGSE---NYK 1846
 Db 1754 -----IDTDAIQGNESCDISKHSETTERTPKHVDHQ-----ICDDAQODLNDLDE 1803
 Qy 1847 MIDSGKRN-----KOLDSKAYIVQTHVIPPDEKELOERTEFRSHNIRPFME 1899
 Db 1804 KLTETGAVPTGECSEGEQSPDTNVE-----PREDKTOGSSECSISLSEFSGNALVP 1855
 Qy 1900 MEPTQGRQGVBEQCKRRTILTAIHCFPVYKXIRPVWYJHHTDNPTEVADIDEMSKV 1959
 Db 1856 MDPL-----GNOED-----IH-----NLOLRKETSINENRLTHV-IEDRDKV 1893
 Qy 1960 A-----ELRQCSSAEVDMIKQLKLGVSVOVNAGPLAYARAFLDOTTKRYPDNKVYL 2015
 Db 1894 EBLNEMELDKHLHGEVQWTKTEIACIELEKIYGLKKNSDISE-GLRFPSCDHQL 1952
 Qy 2016 LKEVRQVYACGQALAVN-ERLIKED-----OLEYOEEMKANYREMAKELSEIMEHOICP 2070
 Db 1953 LQRV--ETSEGLNSDLEMHADKSRREDIGDNVAKVNDSWYERFLDVENELSRIR----- 2004
 Qy 2071 LEKTSVLPNSLHI 2084
 Db 2005 -SEKASIEHEALYL 2017

RESULT 5

US-08-508-836A-8
 ; Sequence 8, Application US/0850836A
 ; Patent No. 5777093
 ; GENERAL INFORMATION:
 ; APPLICANT: Shlloh, Yosef
 ; APPLICANT: Taglio, Danilo A.
 ; APPLICANT: Collins, Francis S.
 ; TITLE OF INVENTION: Ataxia-Telanglectasia Gene
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Reising, Echington, Barnard & Perry
 ; STREET: P. O. Box 4390
 ; CITY: Troy
 ; STATE: Michigan
 ; COUNTRY: US
 ; ZIP: 48069
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/508, 836A
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kohn, Kenneth I.
 ; REGISTRATION NUMBER: 30,955
 ; REFERENCE/DOCKET NUMBER: P-313 (TAU)
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (810) 689-3500
 ; TELEFAX: (810) 689-4071
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3056 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-508-836A-8

Query Match 1.6%; Score 177; DB 1; Length 3056;
 Best Local Similarity 17.3%; Pred No. 7e-05;
 Matches 435; Conservative 348; Mismatches 83C; Indels 918; Gaps 118;
 Qy 66 LINDCLEMLLPYDDFQTALILRQGRYICSTVPAKAEERAGSLFVTECIKTYSN-DWHLV 124

Db 271 LINDSLKEVYI--ELFQLOI-----YIHPRKGAATQCKG-----AYESTKRSI 311
 Qy 125 NYKVED-----YSGFEROLPNVYVLDKLPVHYVEYDEVDQEDAAISGS 170
 Db 312 LYNLYDLVNEISHIGSKYSSGFRNIAVENLIELMADICHQV----FREDTRSLFI 366
 Qy 171 QKGGITTKGWLYKGNMNSAIVTWRSFRRRFFHLIQLDGGSYKEFFLD-LQKEPKGSIF 229
 Db 367 SQSYTTQ-----RESSDYVPCRRK-----IELG-----WEVIDHLQKSGNDPL 409
 Qy 220 LGFL-----YGVFRNNKV-----RRPAFLK-----MDKS 256
 Db 410 VPMLOIATOLISKYPASLPNCELSPMLILSOLPQQRHGRTPYVLCLEVALCQDKR 469
 Qy 257 SYLLAADSEVMEETITLNLKILQNPFAAOKENKNGS-----HEDDESK 303
 Db 470 SNLESSQKSDLLKLM--NKIWCITTFGSISSBOQAEFGLLGAIIGSLVEVDREBK 525
 Qy 304 L-EGSG-----SGLDYLPFLAKSAREABIKLKSERYK- 336
 Db 526 LFTSGACRPSRCVACCLTALITISVPGAVMMGIEQNNCEYNRS-----PSLK-ESIMKW 579
 Qy 337 -LFP-LDPPAQKLDSSAEPV--KSFEKFGKRIYVKNDLSFNLQCCVA----- 383
 Db 580 LLFYQLBGD--LENSTEVPPLHSNFPHLVLEKILV-----SLTMKCKAMNPFQGV 631
 Qy 384 -----ENBGGPTNNVPFV--TLSLFDI-----KNRKISADFHVDLHNHSV 424
 Db 632 ECEHHKHKEKESISEVELFLQTFDKMDPLTVRECGIERKHOSSIGPSVHONKESLD 691
 Qy 425 ROMIATTSPALMGSGPETOS-----ALRGILHR-----A 454
 Db 692 RCLGLSGLQNLNNTSSRTNSETLVRCGRLLVGVLCGCTWGVIAEBAAYKSELFOKANS 751
 Qy 455 AMQYPRKQIFSVTCPHDIFLVARIKVLGSIHCAEPYKSSDSVA-----Q 505
 Db 752 LMQAGESITLPRKNTNEPRIGSLRNMWQ-LCTRCLSNCTKKS-PNKIAGPFLRLTSS 809
 Qy 506 KYLKNAKACQRLQY-RMPF-AMAAATLFPDASGNLDKNAFSAIYQDSNKLSND--- 560
 Db 810 KLMDIADI CKSLASFIRKPPDRGEVESMEDTNGNL-----MEVEDOSSNNLFDNDPD 863
 Qy 561 -----DMLKLP-----ADPRKEK 574
 Db 864 SSVSDANPESGOSTIGAINFLAEVYLSKOLLPLDMLKFLCLCTVTTQNTTVSFRADI 923
 Qy 575 MAKLPVILGNLDITIDNVSDFPYVNSSYIPTRQFETCSKTPITFE-----VEEF 625
 Db 924 RKLLMLID-----SSTLEPTSLHLMYLMKLKELPGSEYPLPMBDV 966
 Qy 626 VPCIPKHTQPTTY-----TNHLVYTPKYL--KYDSQKS----- 657
 Db 967 LELKPLSNVCSLYRRDQCKTILNHVAVKMLGOSNMSEENTRDQOGLTVIGAFW 1026
 Qy 658 -----PAKANIAICB-FKSDSEDSQPLKCIYGR--PGGVPTBSAFALVHH 705
 Db 1027 HLTERRKTIYSVRNALVCLTLLBADYSKRALINWVGKDFPVNSVFTQ--FLDNNHQ 1084
 Qy 706 -----ONPEFY 711
 Db 1085 VRMLAASINRLFDOTKDSRLKALPLKLOTAFAENAYLKAQGMREMSHASNPEFL 1144
 Qy 712 DEIKLELPTQLHEKHHLLTFPHV--SGDNSKSKSTKRDVVEYQVGSWMLPKDQGVV 769
 Db 1145 DEI-----YNRKSVLLTILIAVVLSCS-----PICEKQALFALCKSVENGL- 1185
 Qy 770 TSEOHIPVSA-NLPSGYLYQDELG--MGRHYG-PEIKWVDGKPLKISTH-LVSTVYTO 824
 Db 1186 --EHLVAVKULEKYSSETFGYRLDFMASHLDIVLEKLNQDTBNYLNSSPFIILANTT- 1242
 Qy 825 DQHLNFPQYQCKT-----ESGAQALGNELVYKLSLHA-MEGHVWIAELPTLI 873

```

Db 1243 --NIEDFYRSCYKLLIPLHVLIRSHFDEVSIANQIOEDWKSLLTDCFPKILVNLIFPAY 1300
Qy 874 QLFVRLTRATOEAVY-----NVTYIIVHVAOCHEEGES-- 909
Db 1301 EGTROSGMAOQRETAKKYDMLKSENLOKJOIDHFTSNLPBVLVLEMLTLEHPANSSAS 1360
Qy 910 -----HLRSYKAYKABPYVASEYKTHEBELTKSMITLKESADP 950
Db 1361 QSTDLCDGSGDLDPAPNPPHPSHVIAKATFA--YISNCHKT-----KLKSLILELSKSP- 1413
Qy 951 LHSNKLKTSWFFPDVLKISMAOHLIENSKVLLRNQRPASVYHAAVEVYVNMAMPHIQ 1010
Db 1414 -----SYQILALICEOAAETNNV--YKGRILIKIYH-----LFVSLIKDI- 1453
Qy 1011 KFRDNPASKNANHSLAVFIKCGFTFMDGFVKOJNNYISCPABGDPKTLREYKFEPL 1070
Db 1454 -----KSGIGAMAFVLRDVITYT-----LIHYINQRPSCIMDVLSLS--FSLCCDLS 1499
Qy 1071 VVCNHEHYIPLNLPMPFGKRIQRYODLOLDYSLTDEFCRNHPLVGLLEEVGTALQEBR 1130
Db 1500 QVC-----QTAVTYCKDALENHL-----HYIVGTLL-----PLVYEQV 1532
Qy 1131 EYRLAISVYKULLIHSFPDDRYSRSHQARIATLYPLFGLLIENVORINVRDVSPPV 1190
Db 1533 EYQKQVLDLKYLVINDKNE-----NLV-----ITIKLDPFP- 1566
Qy 1191 NGMTYKDESLALPAPNPLVTPQKSTLDSLHKDGLGAI SGIASPYTTSTPNINSVRA 1250
Db 1567 -----DHVVEKDL-----RITQOKIK 1582
Qy 1251 DSRGSLISTDSGNSLPERNSEKSNLSLKHQOSETLGSVVRCDKLDQSEIKSL--LMCP- 1308
Db 1583 YSRGPR-----SLBEINHPLSVS-----YDALPLVTLLEGKDLRQL 1621
Qy 1309 YLKSNSDALFTYMKKASTSELMDFITSEVCLHOFQYMKRYIARTG-----MHAR 1362
Db 1622 ELHKQOMVDIM-----RASODNPQODIMVLYV--NLQLSKAIIHHTGEKEVLEAVGSC 1674
Qy 1363 LQOLGSLD--NSLTFNHSYGHSDADVLHQSILLEANIN--TEVCLTAL--DTL-----SL 1410
Db 1675 LGEVGIIDSTIAIHSKQASATYKAL--KLEDEKELQMTFIMLYTNLTVLEDQVAVRA 1732
Qy 1411 FTLAFKQULADHGHN--PLMKVFPDVLCLFQKHQOSETA-----LKNVFTALRSL 1459
Db 1733 AVTCLKNILATKTHGSHFWEIYKATTDPMLAYLOPRTSRKKFLVPRPKENPBEFLDII 1792
Qy 1460 -----IYFESTFEGRDMCALCYELK--CNSKUSITEASQLIYFLM 1505
Db 1793 NMIPLSENHDIWIKTLTCAFLDSGCKEIL--QLKPMCEVKTDFCQTVLPYLHIDL 1850
Qy 1506 RANFDYTKKSPFRTHLOVYIISQALAVVIGIGTRFQOOSLIINNCANSRLIKHTSF 1565
Db 1851 LODTNESW--RNLISHTVQGFPT-----SCLHFSQTSHTTPPAL 1889
Qy 1566 SSDVKD-----LTKRIITVLMATAQKHEHNPENMLVDQYSLAKSYASTPELAKTULD- 1619
Db 1890 DSESEHFFRCJLDKSKQRTMLAVVDIMRQKRS-----SGTIFNDAMFWLD 1936
Qy 1620 ---SMARIHVKGDLSEAMCYVHTALV--AEYLTRKEAVQWEPRLPHSHSACLRRSG 1675
Db 1937 NYLEVAKV-----AOSCAHFTALLYAEIYADKKMDQO-----EKS 1974
Qy 1676 GVPROGCTAFVITPNIDEASMBDVGMODVHFNBDVLMEL-----LEQCADGLM 1726
Db 1975 LAFEBGOSTTI--SSLSEKSEETGISIQ-----DLLEIYRSIGEPDLSGCGG-- 2024
Qy 1727 KAERYELADIYKLLIPIYEKRRDPERLAHLYDTLHRAVSKYEVHSGRL----- 1778
Db 2025 -----KMLQPI--TRLRTYEHA--MMKCALVTYDLERAISSSRQAGIIQALON 2070
Qy 1779 LGTYFVAFFGGAQYQFTDSETDV-----GFEDDEGKEVIYKPK 1821
Db 2071 LGLCHILSYVLKGLDYENKDWCPLEBELHYOAMRMQMCHDCTSVSKVEGTS--YHESL 2128

```

```

Qy 1822 LTPLEISORLLKYSDKFGSENNYKMODSGVNPKDLSKYAVI----- 1866
Db 2129 YNALOSLADREHSTYESLTKARKEVEBMCK---RSLSEYYSLYPTLSRQAI GELESI 2185
Qy 1867 -----QVTHVIPFEDEKELOERKTEPERSHNIR--FWEMPTQTKRGQVEBOCK 1917
Db 2186 GELFSRSYTH-----RQSEVYIKQKHQSLKDSDFSFQEPFIMAL----- 2226
Qy 1918 RTITLTHCPFYVKKRIPWYQHHTDNLPIEVALIDSKVYABLRQCSSAEVDMIK-- 1975
Db 2227 RTVI-----LEILMEKEMDNSQRECIKIDILTKLVELSTIARTPKNTOLPR 2273
Qy 1976 --LQKLGSVSVQVNAGPLAVARAF-----LDDTNTKRYPDNRYK 2014
Db 2274 AIFQIKQVNSVSCGSEMQLEAQVFMAKESQSLASTLKQMIKLDLSCAANPS----- 2329
Qy 2015 LKEVFRQFVEACGALA-----VNERLIKEDOLEYQOEBKANY--REMAKEL 2060
Db 2330 -LKLTYTECLRVCGNWLAEFTCLBNPAVINQTYLEKAVENAGVDESSDEL 2379

```

RESULT 6

US-08-629-001A-3

Sequence 3, Application US/08629001A

Patent No. 5858661

GENERAL INFORMATION:

APPLICANT: Shiloh, Yosef

TITLE OF INVENTION: ATAXIA-TELANGIECTASIA GENE AND ITS

TITLE OF INVENTION: GENOMIC ORGANIZATION

NUMBER OF SEQUENCES: 139

CORRESPONDENCE ADDRESS:

ADDRESSEE: Kohn & Associates

STREET: 30500 No. 5858661thwestern Hwy.

CITY: Farmington Hills

STATE: Michigan

COUNTRY: US

ZIP: 48334

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/629,001A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kohn, Kenneth I.

REGISTRATION NUMBER: 30,955

REFERENCE/DOCKET NUMBER: 2290,00032

TELECOMMUNICATION INFORMATION:

TELEPHONE: (810) 539-5050

TELEFAX: (810) 539-5050

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 3056 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

US-08-629-001A-3

Query Match 1.6%; Score 177; DB 2; Length 3056;

Best Local Similarity 17.3%; Pred. No. 7e-05;

Matches 435; Conservative 328; Mismatches 830; Indels 918; Gaps 118;

```

Qy 66 LINDCIREMLLPYDDFOGAILLRQGRYICSTVPAKAEBAOSLPYTBCKITKYN--DWHLV 124
Db 271 LINDSLKEVII--ELFQLOI-----YIHPKAGATTOEKG-----AESTKMSI 311

```


QY 125 NKYED-----YSGERQLENKVKLDKLFVHYEVEDEVEDKEDDASLGS 170
 Db 312 LYNLDVLVNEISHIGSRKSGSGFNIAVKENLIELMADICHQV-----FNEEDRSLERI 366
 QY 171 OKSGITTKGMLYKGMNNSAIVTMSFKRPFHILQIGDGYKPELKD-LCKEPRGSIF 229
 Db 367 SOSYTTT-----RESSDYVPCRRKK-----IELG-----WEYIKDHLKQSONDFL 409
 QY 230 LGFL-----YGVFPNNKV-----KFAFELK-----MODKS 256
 Db 410 VPMLOIAQOLISKTPASLPNCESLPMLTISQLLPQQRHGERTPYLRCLTEVALCODR 469
 QY 257 SYLLAASEVEWEWITTLINKILQLEFAMOEKNGDS-----HEDEOSK 303
 Db 470 SNLESOKSDILKLM-----NKIMCTFRGISSEQKQENFCLGAIQISLVEVREFPK 525
 QY 304 L-EGSG-----SGLSYLPBLAKSAREAIKUKSESRYV- 336
 Db 536 LFTGSACRSCPAVCCLTALTTSIVPGAIVKMGIEQNMCEVNS-----PSLK-BSIMKM 579
 QY 337 -L-FY-LDDPAQKLDPSASAEV--KSFEKFGKRIYKCNELSFNLQCCVA----- 383
 Db 580 LIFPYOLEBG--LENSTEVRPILHSNPHLVLEKILV-----SLTMKCKAMNPROQSV 631
 QY 384 -----ENEGPPTTVNPPV--TSLFDI-----KYNRKISADPHVDLHNSV 424
 Db 632 BCEHHKKEBELSFSEVEBELPLOTTPDKMDPLIYRECIEKQSSIGFSVHQNLKESID 691
 QY 425 ROMIATTSBALMNGSGPETS-----ALRGILH-----A 454
 Db 632 RCLGLISEQLNNYSEITNSEITLVRCSHLVGLVGCYCVGVIAEENYKSELFOKANS 751
 QY 455 AMQYKQGISFYTCPHDIFLVARIEKVLQSGIITHCAEPMKSSDSKYA-----Q 505
 Db 752 LMQCGESITTEKONKTEEFRIQSLRMMQ-LCTRCLSNCTKKS-PNKIASGFLRLTIS 809
 QY 506 KVLKNAKQACQRLGOY-RMP-AMARILFKDASGNLDKNAHPAISLYRODSNKLSD-- 560
 Db 810 KLMNDIADICKSLASFIKKPFDREGEVSEMEDTNGNL-----MEVEDQSSMNLFENDYD 863
 QY 561 -----DMKLCL-----ADPRKPK 574
 Db 864 SSVSDANEBGESOSTIGALINPLAEYLSKODLLFLDMLKFLCLCVTTAQTNTVSPFAADI 923
 QY 575 MAKLPVILGNLIDITIDVNSDFPNYVNSGYIPTKQETCSKTIPRE-----VEEP 625
 Db 924 RKKLMLID-----STLEPTKSLHLMYLMMLKELRGEYPLMEDV 966
 QY 626 VBCIPKHTQPYTY-----TNHLYVPRYL-KYDSQKS----- 657
 Db 967 IELLKPLSNVCSLYRRDOVCKTILNHVLHVKNLQGSNMDSBENTRDAQOGLTVIAGAM 1026
 QY 658 -----FAKARINAIQIE-FKDSDEEDSOPKICTYR--PGGVYFTRSAPAVLHNN 705
 Db 1027 HLTTERKXIFSVRMALVNLCTLLLEADPYSKWAILNMKGDFPVEVFTQ--FLADNHQ 1084
 QY 706 -----ONPEFY 711
 Db 1085 VAMLAESTINRLPOTKGDSSRLKRLPYLKOQTAEBNAYLKAQSBEMKMSAENPEYL 1144
 QY 712 DEIKIELPTOLHEKHLNLTFFPHV--SCNNSKSGTKKRDVETQVGYSMPLKDXGRV 769
 Db 1145 DEI-----YNRKSVLLTLIAVVLSCS-----PICBKQALPALKSVKENG- 1185
 QY 770 TSEOHIPVSA-NLPSGYLGYOELG--MGRHYG-PEIKWIDGSKPLIKISTH-LVSTVITYQ 824
 Db 1186 --BPHLVKVLKSVETFGYRRLLEDFMASHLDYLVLEMNLQDTEYNLSPPFILLANTY- 1242
 QY 825 DOHLNFPQYCOQT-----ESGAQALGNELVKYLKSLHA-MEGHMTIAFLPTILN 873
 Db 1243 --NIEDFYASCYKVLPHLVISHRDEYVSIANQIOEDWKSLLTDCFPKILVNLDPYRY 1300
 QY 874 QLFVLTTRATOREVAN-----NVRVLIHVVAQCHEEGLES-- 909

Db 1301 BGRTRSGNAQORETATKYVMDLKSENLGKOIDHLPISNLPEIIVELMTLHEPANSSAS 1360
 QY 910 -----HLRSYKAVYKAEPPYASRYVHEBELTSMITLKPSADF 950
 Db 1361 OSTDLCDSSGDLDPANPPHPSHVITKATFA--YISNCHKT-----KLSILEILSKSPD- 1413
 QY 951 LTSNKLTKYSWFFPDVLIKSMAOHLIENSKYKLJRNORPPASYNHVAVEVVMMLPHITQ 1010
 Db 1414 -----SYCKILAIQEOAALETNNV--YKGRHILKITYH-----LPVSLILKDI-- 1453
 QY 1011 KFRNPESKXANNSLAVFIRKCTFTMDRGVFKOINNYISCPAGDKTLPEYFEBLR 1070
 Db 1454 -----KSGLGAMAFVLADVIYT--LIHYNORPCSIDMVSJRS-FSLCCDILS 1499
 QY 1071 VYCNHEHYPLNLPMPFGKRIORYQDLODYSLDBCRNHPFLVGLLREVTALQER 1130
 Db 1500 QVC-----OTATYCKDLLENHL-----HVIYGLI-----PLVYEOV 1532
 QY 1131 EYRLIAISVNLKNLIKSPDRYASRSHQARIATLYLPFGILLIENVCRIWVDSPEPV 1190
 Db 1533 EVQKQVLDLKLKYLVIDNKDNE-----NLV-----ITIKLDPFP- 1566
 QY 1191 NAGMTVQESIALPANDPLVTPQKSTIDNSLHKDLGAISGIASPTTSTPNINSVNA 1250
 Db 1567 -----DHVVFQDL-----RITQOKIK 1582
 QY 1251 DSRGSLISTDSGNSLPERNSEKSNLGDHQSSITGNSVNRCDKLDQSBISL--LMQFL 1308
 Db 1583 YSRGPF-----SLLEEINHLISVS-----YDALPRLRLBGLKDLRRQL 1621
 QY 1309 YILKMSDDELFTYNNKASTSELMDFTTISEVCLHQPOYMGKRYIARTG-----MMHAR 1362
 Db 1622 BLHKQWMDIM-----RASQONPDQGIWVKLV--NLQLSGMAINHTGKEXVLEAVASC 1674
 QY 1363 LQQLGSLD-NSLTNHSYGHSDADVLHOSLLEBANTA--TEVCLTAL-DTL-----SL 1410
 Db 1675 LGVEVPIPFSTIALQSHSDASTTKAL--KLEDEKELQWTFIMLTLYINNTLVEDCVKVSBA 1732
 QY 1411 FTIAFKNOLLADHGN--PLMKXVEDVYLCLFQKQKQSTA-----LKNVFTALSL 1459
 Db 1723 AVTCLKNILATKGTGSPWEIYKMTTDPMLAYLQRPRTSRKKFLVBRDKENPFGLDI 1792
 QY 1460 -----YKPFSTYEGRADMCALCYEILK-CCNSKLSIRTEASQULYFLM 1505
 Db 1793 NLMITLSENHDIWITLTCALFDSGTCCEL--QLKRPKEVKTDPFGQVLPYLIHDL 1850
 QY 1506 RNNFDYTKSKSFVRTHLOVISVSQLADLVVIGGSTRFOQSLIINNCAUSDRLIKHTSF 1565
 Db 1851 LQDTNBSW-RMLSTHYQGFT-----SCURHFSQTSRSTTPANL 1889
 QY 1566 SSDVD-----LTRKIRVLMATAQMKENHENDREMLVDLOYSLAKSYASTPRLKTYLD 1619
 Db 1890 DBESHFRCCIDKKSQRTKLAUVDMRQRPS-----SGTIFNDAPWLDL 1936
 QY 1620 ---SMARLHVNGGDLSEBAMCVNHTALV-AEYLTRKCAVOWPEPRLPHSHSACURRSRG 1675
 Db 1937 NYLBAKY-----AOSCAHFTALLYAEIYADKSMDOQ-----EKRS 1974
 QY 1676 GVFRQSGTAFRVTIPNIDEASMMEDVGMQVHNEVDLMBL-----LEOCADGLW 1726
 Db 1975 LAFEBSQSTYI--SLSEKSEKEETGISLQ-----DILLBIYRSIGRPSLYCGCGG-- 2024
 QY 1727 KAERYELADIYKULPIYERKRDPERLAHLYDTLHRAYSKVTYEMHSGRL----- 1778
 Db 2025 -----KMLQPI-TRLRTYENHA--WMGKALVTYUDETALPSSTRQAGIITQALON 2070
 QY 1779 LGTYFRVAFPGQAQOYQFTDSETV-----GFFEDBDGKRYIYKPEK 1821
 Db 2071 LGLCHILISVYLKGDYENKDCPBELEHLYQAARNMOMDCTSVSKVBGTS--YHESL 2128
 QY 1822 LTPLEISQRLIKLYSDRFSGSENVMMIDQSKVNEPKDLDSTRYANI----- 1866

Db 2129 YNALGSLDRBEPSTFESYLKARVKEVEBCK---SLESVSYLPTLSRLQAGLEBGI 2185
 Qy 1867 -----QTHVLPFDEKELOERKTEPERSHNIR---FMFEMPFTQTKRGQVGEQCK 1917
 Db 2186 GELFSSSVTH-----RQSEVYIKQKQSOLKSDSDFSEPPIAL----- 2226
 Qy 1918 RTILTALHCFPPYKRIIPVMOHTDNLPIEVALIDEMSKVAELRQCSAEVDMIK-- 1975
 Db 2227 -RTVI-----LEIMEKEMDNSQRECIKIDILTKHLVELSLARTFKNTQLPBR 2273
 Qy 1976 --LQLLQGSYSQVNAAGPLAVARAF-----LDDNTKRYPNNKKA 2014
 Db 2274 AIFQIKQVNSVSGVSEWOLEEQAQVFAWAKEQSLALSLIKOMIKKLIDASCANNS-- 2329
 Qy 2015 LKLEVPFQVFAACGQALA-----VNERLIKEDOLEYOEMKANY-REMAKEL 2060
 Db 2330 -LKTYTECLRVCGNMWLAETCLENPVAVIMQTYIEKAVAVAGNYDGESSDEL 2379

RESULT 7
 US-08-642-274D-3
 / Sequence 3, Application US/08642274D
 / Patent No. 6200749
 / GENERAL INFORMATION:
 / APPLICANT: Shiloah, Yosef
 / TITLE OF INVENTION: MUTATED FORMS OF THE ATAXIA-TELANGIECTASIA GENE AND METHOD TO
 / FILE REFERENCE: 22900003
 / CURRENT APPLICATION NUMBER: US/08/642,274D
 / NUMBER OF SEQ ID NOS: 220
 / SOFTWARE: PatentIn Ver. 2.1
 / SEQ ID NO 3
 / LENGTH: 3056
 / TYPE: prt
 / ORGANISM: Homo sapiens
 US-08-642-274D-3

Query Match 1.6%; Score 177; DB 4; Length 3056;
 Best Local Similarity 17.3%; Pred. No. 76-05;
 Matches 435; Conservative 328; Mismatches 830; Indels 918; Gaps 118;

Qy 66 LNDCLREMLLPYDDFOTAILRRQGYICSTVPAKAEEAQSLEFTECIKTYS--DWHLV 124
 Db 271 LNDLSLEVIIL--ELFOLQI-----YHHPKGAKTQEGK-----AYESTKRSI 311
 Qy 125 NKRYED-----YSGFERQLNKKYVKDKLPVHYVEVDEVDKEDDASLGS 170
 Db 312 LYNLYDLVNEISHIGSRGKYSGFNNIAVKELIEIMADICHQV-----FNEDETSLEI 366
 Qy 171 OKGGLTKHGMLYKGNNSAISVTMRSPKRRFPHILOLGDGSYKFEFLKD-LQKEPKSIF 229
 Db 367 SOSYTTTQ-----RESSDYVPCRRKK-----IELG-----WEYIKHLQKSDNPL 409
 Qy 230 LGFL-----YGVSPNNKV-----RRFAFLK-----MODKS 256
 Db 410 VEWLQIATOLISKYPASLPNCELSPLMLISQLPQORGERPTPYLRCLTEVALCODKR 469
 Qy 257 SYLLADSEVEHEBETIINKILQNFELAMOEKRGDS-----HEDDEQK 303
 Db 470 SNLESSQKDLKLM--NKIWCFIFRGISSEQKAEFGLGAILQISLVEVDREFFK 525
 Qy 304 L-EGSG-----SGLDSEYLPFLAKSAREAEIKLKSEGRVK- 336
 Db 526 LFTGSAICRSCRAVCCLTALTTSIVPGAVKGIENMCEVNS-----FSLK-BSIMKA 579
 Qy 337 -LPY-LDPPAOKLDFSSAPEV--KSPFEKFGRIILVKNDLSFNLQCCVA----- 383
 Db 580 LIFQYLEGD--LENSTEVPRILHSNFPHLVLEKILV-----SLTKNCKKAMNPFQSV 631
 Qy 384 -----ENBEGPTNVPEFPV--TLSLFDI-----KYNKIKISADPHVDLHNSV 424
 Db 632 ECEHHKKEELSFSEVEBELFLOTTFDKMDFLIYVREGCIEKQSGISGVHONKESLD 691

Qy 425 ROMIATSPALMNGSGPETOS-----ALRGILHE-----A 454
 Db 632 RCLLQSLQNLNNYSSEITNSETLVRCGRLLVGVLCGYCNGVLAIEEBAAYSELFQKANS 751
 Qy 455 AMQYKQGI FSVYTCCHPDI FLVARI EKVLQGSITHCAPYKMSDSSKYA-----Q 505
 Db 752 LMQAGESIITLPKNKTNEEFRIKSLRNMQ--LCRCLSNCTKKS--PNKIASGFPLRLTS 809
 Qy 506 KVLKNAQACORLAGY--RMPF-AMARTLPDASGNLKNARFSAIYHODSKLSD--- 560
 Db 810 KLMDIADI CKSLASFIKKPPDRGEVESMEDDTGNL-----MEVEDOSSNNLNDYD 863
 Qy 561 -----DMUKL-----ADFRKEK 574
 Db 864 SSVSDANPEGSGSTIGAINPLAEEYLSKODLLFMDLKFLCLCVTAQTNTVSRADI 923
 Qy 575 MAKLEVILGNLDTITDNVSSDFPNVNSSYLPKQFETCSKPTTFE-----VEEF 625
 Db 924 RKLLMLID-----SSTLEPTKSLHLMYLMKLPLGEEYPLPMBDV 966
 Qy 626 VPCIPKHTQPYTY-----TNHLYVPKYV--KYDSQKS----- 657
 Db 967 LELKPLSNVCSLYRRDQVCKTILNHVLHVVKNLQGSNMDSBENTRDAQOQFLTYIGAFW 1026
 Qy 658 -----FAKARNIAICIE-FKDSDEDSQPLKCIYGR--PGSPVETRSAPFAVLHH 705
 Db 1027 HLTFRKXIFSVRMALVNCIKTLLEADPYSKMALINWGXKQFPVNEVFTQ--FLADNHQ 1084
 Qy 706 -----ONPEFY 711
 Db 1085 VMLAEBISINRLFODTKDSSRLKALPLKLOQTAFENAYLKAQREMSHSAENPETL 1144
 Qy 712 DEIKTELPTQLEKHHLLTFPHV--SCDSSKSTKRDVETPOGVSWLPLDGRV 769
 Db 1145 DEI-----YNRSAVLTLIAVLSCS-----PICEKALFALCKSVENG-- 1185
 Qy 770 TSECHI PVSA-NLPESGYLQYELG--MGRHYG-PEIKWVGGKPLKLISTH-LVSTVYV 824
 Db 1186 --EPLLVKLVLEKSETGYRLBDFMASHDYLVLEMLNODPEYNLSSPFLILANT- 1242
 Qy 825 DOHLNFPQYCOKT-----ESGAOLGNELVKYLKSLHA-MEGHVMIAFLPTLN 873
 Db 1243 --NIEDFYRSCYKVLPIHLVIRSHPEVKSIAHQIOEBMSILTDCEPKILVINILPYAY 1300
 Qy 874 QLFVLTATQEEVAV-----NVRVLIHVAQCEBLES-- 909
 Db 1301 BSTRDGAQOREIATKYVDMKSENLLGKOIDHLFISNLPETIVELMTLHEPANSAS 1360
 Qy 910 -----HLRSYKVAVKABPVYASEYKTVHEELTKSMTTLKPSADF 950
 Db 1361 QSTDLCDPSGDDPAPNPHPSHYIKATPA--YISNCHK--KLKSILEILKSPD- 1413
 Qy 951 LTSNKLKYSWFFPVLIKSMAOHLIENSKYKLYLRNQPSPSYHNAVETVYNNMLPHITQ 1010
 Db 1414 -----SYOKILAIACEQALETNNV--YKGRHILKIYH--LPVSLKLDI-- 1453
 Qy 1011 KFRONPEASKNANSHLAVIFKRCFTFEMDRGVFQOINNYISCFARPDKPTJEYFEFLR 1070
 Db 1454 -----KSGLGAMAFVLRDVIYV--LIHYINRPSCLINDVLSRS--PSLCCDLIS 1499
 Qy 1071 VCNHEHYIPLNLPMPFGKRIORQDQLDYSLTDFECRNHFLVGLLRLREYGALEFR 1130
 Db 1500 QVC-----QTAVTYCKDALENHL-----HVIYGTIL-----PLVYQOV 1532
 Qy 1131 EYRLAIVLKNLLIKSHFDDRYSRSHQARIATLYLPGLLLENVORINVRDVSPPV 1190
 Db 1533 EYQGVLDLKLKYLVIDKNDE-----NLV-----ITIKLDPP- 1566
 Qy 1191 NAGMTVKBESLAPRVNPLVTPQKSTLDNSLHLDLGAISGASPYTTSTPNNINSVNA 1250
 Db 1567 -----DHVVPKDL-----RITQOKIK 1582

```

OY 1251 DSRSLISTBDCGNSLSPENSSKNSLDDGQSSGLTNGSVNCDKLDQSEISL--LMCF 1308
Db 1583 YSRGPF-----SLIEELNHFLSVS-----YDALPLTRBGLDKDLRQL 1621
OY 1309 YILKSMSDDAFTYWNKASTSELDNPFPISEVCLHOFQYMKRYIARTG-----MMAR 1362
Db 1622 ELKHDDQWDM-----RASQONPDQGIWKLTV--NLLQLKMAINHGEKEVLEAVGSC 1674
OY 1363 LQQLGSLD-NSLTNRHSGHSDADVLHOSLLEANIA--TEVCLTAL-DTL-----SL 1410
Db 1675 LGEVGPIDFSTIALQHSKDASTYAL--KLPEDEKELQWFTMLTYLNNLTVEDECVKRSA 1732
OY 1411 FTLAFKQDLADHGN--PLMKKVPDYVLGCLQKHGSTA-----LKNVFTALBSL 1459
Db 1733 AVTLCKNLATKTGHSFWEIYKMTTDEMLAYLQFPRTSRKIFLEVPREDKENPEGLD 1792
OY 1460 -----LYKPFSTFEGRADWCAALCEYELK-CGNSKLSIRTEASOLLYELM 1505
Db 1793 NLMTPLSENHDIWIKTLGALDGGGTCCLL--QLKPKMEVUTDPCQTLPLVINDIL 1850
OY 1506 RNNPDYGGKSFVATHLOVITISVSQLADVVGIGTRFQGLSLIINNCANDRLIKHTSF 1565
Db 1851 LQDNTESW-RMLSTHVGGFTL-----SCLRHSQSRSTTPANL 1889
OY 1566 SSDVKD-----LTRIRIVLMAIQAOMKENHENDREMLVDLOGLSLAKSVASTELAKTMD 1619
Db 1890 DSEEHFRCCLDKKSQRTMLAVUDYMRQRPS-----SGTIFNDAPWIDL 1936
OY 1620 ---SMARIHVKNQDLEBAMCVVHATLV--AEVLTREKAVQWEPYLPHSACILRSRG 1675
Db 1937 NYLEVAKV-----AQSCAHNTALLYHEIYADKSSMDQ-----EGRS 1974
OY 1676 GVFRQCGTAFRVITPNIDEBASMMEDVGMQDVHFNEDVLMHL-----LEQCADGLW 1726
Db 1975 LAPEGSGSTTI--SLSEKSEKEBTGISLQ-----DLTLHYRISGEPDLSYGGG-- 2024
OY 1727 KAEKYEELIADLYKLIPIYKRRPERLAHLVDTLHRAYSIVTEYMHSGRL----- 1778
Db 2025 -----KMLQPI-TRLRTYEHEA--WKGALVTYDEETALPSSSTRQAGIIQALON 2070
OY 1779 LGTYFRVAFGQAOYQCFETDSETOVE-----GPFEDDGKEXLYKEPK 1821
Db 2071 LGLCHILSVLYKGLDYENKDMCPLEELHYQAMRNMMQDICTSVSKVEGTS--YHSL 2128
OY 1822 LTPLEISQRLLYKUSDKFGSENNVMIODSGVMPKDLDSICAYI-----1866
Db 2129 YNALQSLADREPSFYTESLKYARVAGEVBEMCK--RSLSESYSLPFLTSLRLQATGELESI 2185
OY 1867 -----QVTHVITPFFDEKELOERKTEFERSHNIR--FWMEPMTQTGGKQGVBEQCK 1917
Db 2186 GELPSRSYTH-----RQLEBVYIKMKGHQLKDSDFSIQEPFIMAL-----2236
OY 1918 RRTILTAHCFPVYKGRIPWVQOHTNDLPIEVALDEMSKYAEIROLCSSEAYDMIK-- 1975
Db 2227 RTYI-----LEIIMEXKMDSQRECIKDIIIFKHLVELSILARTFKTOLPER 2273
OY 1976 --LQQLGQSVSVQVNAAGPLAYARAF-----LDDNTKRYVDNKKV 2014
Db 2274 AIFQIKQVNSVSCGVSEWQLEBAQVFMAKKQOSLALSLIKMIKLDLSCAANNPS----- 2329
OY 2015 LKKEVROFVEACGQALA-----VNERLIKEDOLEYOEEMKANY--REMAKEL 2060
Db 2330 -LKLVTYECTLRVCGNMMLAETCLEENPAVIMQVYLEKAVAEVANAVIDESSEDEL 2379

```

TITLE OF INVENTION: ATRX1A-TELANGELECTASIA GENE
 NUMBER OF SEQUENCES: 24
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Kohn & Associates
 STREET: 30500 N. 6211336thwestern Hwy., Suite 410
 CITY: Farmington Hills
 STATE: Michigan
 COUNTRY: U.S.
 ZIP: 48334
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DAT#: _____
 APPLICATION NUMBER: US/08/952.127
 FILING DATE: _____
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Kohn, Kenneth I.
 REGISTRATION NUMBER: 30,995
 REFERENCE/DOCKET NUMBER: 2290,00029
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 810-539-5050
 TELEFAX: 810-539-5055
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3056 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE: _____
 ORGANISM: Homo sapiens
 JS-08-952-127-3

Query Match 1.6%; Score 177; DB 4; Length 3056;
 Similarity 17.3%; Pred. NC. 7e-05;
 Best Local 435; Conservative 328; Mismatches 830; Indels 918; Gaps 118

```

Oy 66 LNDLCEMLPFYDFOFALILBRQRYCISVPAKAEBAOSLPEVCEIKTVNS--DMLIV 124
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 271 LNDLSLKEVYI---ELFQLOI---YIHFKKAKTQEKG-----AYESTKRSI 311
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 125 NYKYEJ-----YSGEFRQLPNKVVLKDLRYHVEVDEVDKEDDASLGS 170
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 312 LVYLYLVLVNEISHIGSRGKYSSSGFRNVAEMNIELMADICHOV-----FNEBTSLEI 366
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 171 OKGIGIKHMLKKGMMNSAISVTMRSFRRRPFHLIOJGDGSYKREPLCD--LOKEPKSIF 229
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 367 SÖSYTTTO---RESSDYSPCKRKK-----TELD---WEVICKDLÖKSÖNDFL 409
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 230 LGEFL-----YGVSPRNKNV-----RRFAFLK-----MÖKS 256
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 410 VPMLOJATOLISKYPASLPNCELSPLMLISQLPQORHGERTPYVRLCTEVALCQDKR 469
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 257 SYLLADSEVEHEWITLIKLIQNLPEAAQEKRGNS-----HEDDQSK 303
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 470 SNLESSQSKDLKLM---NKIWCITTFRGISSEKQÖKÖNPEGLGAILIÖGSLVEYDRFEWK 525
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 304 L-EGSG-----SGLDSTLPELAKSAREABIKLSKESRYK- 336
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 526 LRFGSAKCRSPCAVCCLLTALTTISVPAVVMGIEÖNNCEVNR-----PSLK-ESIMXV 579
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 337 -LFF-LDPDQÖKLDPSSABPEV--KSPEKFGKRIYKNDLSPNÖCCVA-----383
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 580 LLEFÖLEGD---LENSTEVPRILHSGFHLVLEKILV-----SLTMKACRANNFQSGVP 631
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 384 -----ENEBGPTTNVEPFFV--TLSTLFDI-----KYRKISADPHVDLNFVSV 424
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Db 632 ECHHHKÖKEBLSFSEVELFLQTTFDMDPLTVIREGIEKÖSSIGFVHÖNVLKESLD 691
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
Oy 425 RÖMIATTSPALMNGSGPETÖS-----ALRGILHFE-----A 454
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```


ADDRESSEE: Kohn & Associates
 STREET: 30500 No. 626518thwestern Hwy., Suite 410
 CITY: Farmington Hills
 STATE: Michigan
 COUNTRY: U.S.
 ZIP: 48334
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 OPERATING SYSTEM: IBM PC compatible
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/952,014C
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Kohn, Kenneth I.
 REGISTRATION NUMBER: 30,995
 REFERENCE/DOCKET NUMBER: 2290,00028
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 810-539-5050
 TELEFAX: 810-539-5055
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3056 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 US-08-952-014C-3

Query Match 1.6%; Score 177; DB 4; Length 3056;
 Bees Local Similarity 17.3%; Pred. No. 7e-05; Indels 918; Gaps 118;
 Matches 435; Conservative 388; Mismatches 830;

66 LNDCLREMLPPYDDPQTAIRROGRVYISTVPAKAEENOSLFVTECIKTYNS-DMLV 124
 271 LNDLSKEVIL-ELFQLQI-----YIHHPKAKTQEK3-----AESTKMRRI 311
 135 NYKTED-----YSGEFPQLPNKVVKDKLPVHYVEDEVDKEDASLGS 170
 312 LYNLDLVNVEISHIGSRGKSGGFNIAVKENLIELMADICHOV-----FNEIDTSLEI 366
 171 QXGGITKKGWLYKGNNSAISTVMSFKRFRHLIQLGDSYKFEELKD-LQKEPKGSIF 229
 367 SGGYTTTQ-----RESSDYVPCRRK-----IELG-----WEYIKDLQKSONDFL 409
 230 LGFL-----YGVSPNNKV-----RRFAFLK-----MODKS 256
 410 VEWLQIATQOLISKYRPSLNCESLPLMLISQLLPQGRGERPPTVRLCLTEVALCQDGR 469
 257 SYLLAADSEVEWEWITTIINKILQLNFEAMQEKNGDS-----HEDDEQSK 303
 470 SNLSSQSKDLKLMW-----NKIWCITFRGISSEQKQAEFGALIGSLVEVDPEFK 525
 304 L-EGSG-----SGLDSELELAKSAREATIKLKSERVK- 336
 526 LFTGSACRSPCAVCCLTALTTSIVPAVKMGIEQNMCEVNS-----FSLK-BEIMKM 579
 337 -LFY-LDPPAQKLDFFSAPEV--KSPBEKFGKILVKNCDLSFNLQCCVA----- 383
 580 LLFYQLEGD---LENSTEVPPIIHSNFPILVLEKILV-----SLTMKCKKAMNPFQSPV 631
 384 -----ENEGPTTVNEPPPV--TLSLFDI-----KYNRKISADFHVDLNFHFSV 424
 633 ECEHHKDKKEELSPSEVEELPLQTFPDKMDPLTYVRECGIEKQSSIGFSVHQLKESLD 691
 425 ROMIATTSFALNNGSGERTQS-----NLRGLHF-----A 454
 692 RCLLGLSEQLNNYSSEITNSSETLVRCRLVGLVGLACYCYMGVIAEEYKSELFOKANS 751

455 AMQYPRKGI PSVTCPHDI FLVARIKVLQGSITHCAEPYMKSSDSSTVA-----Q 505
 752 LMOCAGESTILFKNKTNEEFRIQSLRNMQ-LCTRCLSNCTYKS--PNILASGFLRLITS 809
 506 KYLKNKAKQACORLQY-RMPF-AMAAKTLFPDASGNLDKNARFSAIYQDSKLSND--- 560
 810 KLMNDIADI CYSLASFIKKPDRGEVESMEDTGNL-----MEVDQSSNNLFRNDYPD 863
 561 -----DMKLK-----ADPRKPEK 574
 864 SSVSDANBESQSTIGAINPLAEYLSKODLFLDMLKFLCLCTTAQTNVSRADI 923
 575 MAKLPVILGNLDITIDNVSSDFPNYVNSYIPTKQFCTSKPTTFE-----VEEF 625
 924 RKKLMLID-----SSTLEPTKSLHMYLMLLKEJLGEERYPLMEBV 966
 626 VECIRKHPOTTY-----TNLYVYPKYL---KYDSQKS----- 657
 967 LELKPLSNVCSLYRDOVDCKTILNHLVHVKNLGQSNMSENRDAGQOFLTVIGAFW 1026
 658 -----FAKARVIAICIE-FKDSDEDSQPLKCIYGR--PGGPVFTSAFAVLHH 705
 1027 HLTERTKTYFSVRNALVNCITLLEADPYSKAILNWKQEPVUNVEFTQ--FLADNHQ 1084
 706 -----ONPEFY 711
 1085 VRMLAASINRLFQDTKGDSRLKALPLKQQTAFENAVYLKAGQREMSHSAENPFL 1144
 712 DEIKLEPLQLHEKHHLLTFPHV--SCDNSKSGTKKRDVETGVGYSMLPLDGRV 769
 1145 DEI-----YNRKSVLLTLIAVLSCS-----PICEKQALPALCKSVKENG- 1185
 770 TSECHIPVSA-NLPSGLYGOELG--MGRHYG-PEIKVWDGKPLKISTH-LVSTVYV 824
 1186 --EPHLVKKVLEKSETGYRLEDFMASHLDYLVLEMLNODTETVNSSPFILLNT- 1242
 825 DQHLNFFOYCOKT-----ESGAQALGNELVLYKLSIHA-MEGHVMIAPLPTLN 873
 1243 --NIEDFRCYKVLPHLVIRSHDFEVASIANQIOEDMKSLTDCPRLVNIILPYFAY 1300
 874 QLFVLTAIQEYAV-----NTRVITHVVAQCEBGLS-- 909
 1301 EGTDRSGMAQOHEFTKYVDMLKSNLLGKOIDHLFISNLPETVIELMTLTHEPANSSAS 1360
 910 -----HLRSYKVAYPEVYASSEKTYHELSMTTILKPSADF 950
 1361 OSTDLCDPSGDDPAPNPPHPSHVIAKTPA--YISNCHKT---KLASILEILKSPD- 1413
 951 LTSNLLKYSWFFPDVLKSMACHLIENTSKVLLNORFPASVYHNAVETVVMMLPHITQ 1010
 1414 -----SYKILALICEQALETNNV--YKGHILKITYH--LFGSLLLKDI-- 1453
 1011 KFRDNPASKNANHSIAVFIKRCFTFMDRGVFKOINNYISCFAPGDKTYLEFYEFELR 1070
 1454 -----KSGLGAMAFVLRDVIYT--LIHYINRPSCIMVSLRS-FSLCCDLIS 1499
 1071 VWCNHEHYIPLNLPMPKGRIRIQRYQDLOVSLDFECRNHPVGLLBRGTALQER 1130
 1500 QVC-----QTAATYCKDALENHL-----HVIQVTLI--PLVEYV 1532
 1131 EVRLIAISVLKNLILKHSFDDRYASRSHQARIATLYLFGILLIKRIVORINRVDSPPV 1190
 1533 EYQKVQVLDLKTIVIDNKNF-----NLY-----ITIKLDPF- 1566
 1191 NAGMTVQDESLALPAVNPVLPQKSTLDNSLHKDLAIGSIASPYTSTPNINSVRNA 1250
 1567 -----DHVVFQDL-----RITQOKTK 1582
 1251 DSRGSLISTDGSNSLPERNSEKSNLGDHQQSSTIGNSVNCDDKJQSEIKSL-LMCTL 1308
 1583 YSRGPF-----SLSEBINHFLSVS-----YDALPLFRLBGLKOLRROL 1621
 1309 YLKSMDDALFTVYNNKASTSELMDFTIISEVCLHQFQYMGKRYIARTG-----MMHAR 1362

```

Db      1622 ELHKDQWVIM-----RASQDNPOQDGMVLYV--NLLOLSKMAINHTEKEVLEAVGSC 1674
Qy      1363 LQQLSLD-NSLTFNHSYGHSDADVLHQSLEANI--TEVCUTAL-DTL-----SL 1410
Db      1675 LGEVGIIDSTAIQHSKASYSYKAL--KLFEDEKQWPFIMLTLYLNTLTVEDCVKVRSA 1732
Qy      1411 FTLLAFNQLADHGH--PLMKVFPVYLCTLOKHQSETA-----LKNVFTALRSU 1459
Db      1733 AVTCLKNILATKTHGFWELKYKTTDPMALYLOPFTSRKKFLVFRPKENFBSLDDI 1792
Qy      1460 -----TYRPFSTFEGRADMCALCYEILK-CNSKLSIRTEASQLLYFLM 1505
Db      1793 NLMIPLSENHDIWIKTLTCAFLDSCGCKEIL--QLKPKCEVKTFCQTVLPYLLHIDL 1850
Qy      1506 RNNFDYTKKSFRTLOVLIISQSLADVIGIGTRFOQSLSIINNCANSDBLIGHTSF 1565
Db      1851 LQDTNBSW--RNLSTHVOGFFT-----SCLNHFQTSSTTPPALU 1889
Qy      1566 SSDVKD-----LTKRIRTVLMATAQMEHENDPEMLVDLOYSLAKSVASTPELRKTWLD- 1619
Db      1890 DSSSEHFRCCLDKKSQRTMLAVDYMRQKRS-----SGTIFMDAFWLD 1936
Qy      1620 ---SMARIHVKGDLSEAMCYVHTALV-AEYLTRKKAQWQRPPLLPHSHSACLRSRG 1675
Db      1937 NYLEVAKV-----AQSCAHFTALYAEIYADKKSMDQ-----EKRS 1974
Qy      1676 GVPFROCTAFRVITPNIDEASMEVDVGNODVHFNVDVMEI-----LEQCADGLM 1726
Db      1975 LAFESQOSTT--SSLSKSKETGSIQ-----DLLEIYRISIGEPDSLGCQGG-- 2024
Qy      1727 KAEVYELADIYKLIPIYEKRDPERLAHLDTLRAYSKYTEVHNSGRLL----- 1778
Db      2025 -----KMLQPI-TRLTVEHEA--MMGKALVTYDLLETALISSRQAGIQTALQN 2070
Qy      1779 LGTFVFAFPFGAAQYQFTDSETDV-----GPFEDDGKEYIYKEPK 1821
Db      2071 LGCHILSYVLKGLDYENKQWCELELHYQAAMRNQMDHCTSVSKEVGTG--YHESL 2128
Qy      1822 LRPFLHSISRLKLYSDKSGSENVKMIQDSKVPNDLDSKYAI----- 1866
Db      2129 YNALQSLRDEFTFYESLKYARVKEVEWCK--NSLSVSYSLYTSLRLQALGELBI 2185
Qy      1867 -----QVTHVLPFPEKELOERKTEPERSHNIR--FMFEMPQTQTKRQGVGEQCK 1917
Db      2186 GELFSFSVTH-----RQISEVYIKMQHSQSLKSDSFOEPHAL----- 2226
Qy      1918 RRTILTALHCFPVYKRIIPVMTQHTDNLPEVAIDEMSKVAELRQLCSSAEVDMIK-- 1975
Db      2227 -RTVI-----LEILMEKEMDNSQRECIKIDILTHLLEVELSILARTFKNQLPBR 2273
Qy      1976 --LQTLQGSVSVQVNAAGPLAVARAF-----LDDNTKRYPNKVK 2014
Db      2274 AIFQIQVNSVSGVSEWOLEEQVFWAKKESGLALSILKOMIKLUDASCAANPS-- 2329
Qy      2015 LKLEVFQVEAGGOLA-----VNERLIKEDOLEYOEEMKANY-REMAKEL 2060
Db      2330 -LKTJYTECLRGVGMWLAETCLENPAINQTYLEKAVAVAGNVGDSSSEL 2379

```

RESULT 10
US-09-360-416-3
Sequence 3, Application US/09360416
Patent No. 6458536

GENERAL INFORMATION:
APPLICANT: Richard A. Gatti
TITLE OF INVENTION: METHODS FOR DETECTION OF ATAXIA
FILE REFERENCE: 510015-222
CURRENT APPLICATION NUMBER: US/09/360,416
CURRENT FILING DATE: 1999-07-23
NUMBER OF SEQ ID NOS: 143
SOFTWARE: FastSeq for Windows Version 3.0

```

; SEQ ID NO 3
; LENGTH: 3057
; TYPE: PR
; ORGANISM: Human
US-09-360-416-3

Query Match      1.6%; Score 173.5; DB 4; Length 3057;
Best Local Similarity 17.4%; Pred. No. 0.00014;
Matches 436; Conservative 330; Mismatches 836; Indels 899; Gaps 119;

Qy      66 LNDCLREMLPFYDFQFALRLRQGRYICSTVPAAEBAQSLFTECKITYNS--DWHLV 124
Db      273 LNDLSKEVYII--ELFQLOI-----YIHPKGAATQORQ-----AYESTKMSI 313
Qy      125 NYKED-----YSGEFPOLPNKVVKDKLPVHYVEYDEVDKEDASIGS 170
Db      314 LYNLYDLVNEISHIGSRKYSGGFRNIAVENLIELMADICHQV-----FREDTRSLBI 368
Qy      171 OKGITTGKWLKYGNNNSAISVTMRSFKRRFPHL1QLDGSYKFFELD-LQKEPKSIF 229
Db      369 SGTSTTQ-----RESSDYVPCRRK-----IEIG-----WEVIDHLQKQNDPDL 411
Qy      230 LGFL-----YGVSPRNKV-----RRPAELK-----MDKS 256
Db      412 VPMLOIATQLSKYPASLPNCESPLMLISQLPQORHGERTPYVLRCTEVALCODKR 471
Qy      257 SYLLAADSEVEMEYITLNLKLOLNFEMAQEKNGDS-----HEDEQSK 303
Db      472 SNLSSQKSDLLKM-----NKIWCITFRGISSEQLOAENFGLLAIIGSLVEVDREPK 527
Qy      304 L-EGSG-----SGLDYLPBLAKSAREAEIKLKSERYK- 336
Db      528 LFTGACRSPCPAVCCLTALTTISVPGAVKMGIGIONCEVRS-----FSLK-ESIMKM 581
Qy      337 -LFF-LDPDAQLDSSAEPV--KSFEKRGKRLYKCNDSFPLQCCA----- 383
Db      582 LFFQLQEGD--LENSTEVPPLIHSNFPHLVLEKILV-----SLTMKCKAAMNFQSV 633
Qy      384 -----ENEGPTTNVPEPFV--TLSPFDI-----KYNRKISADFPVDNHSV 424
Db      634 EENHQKDEBELSFEVEBELFLQTTFPDKMDPLTYREGIEGHOSIGFSVQNLKESID 693
Qy      425 RQMIATTSBALMGSGFETOS-----ALRGILHE-----A 454
Db      694 RCLLGISEQLNNYSSEITNSFTLVRCGLVGVLCYCMGVIAEEBAKSELPQKANS 753
Qy      455 AMQYRKQGISYTCRPHDPLVARIKYLQGSITHCAPYKSSDSKYA-----Q 505
Db      754 LMOCAGESITLFPKNTNEEFRIQSLRNMQ--LCTRCLSNCTKKS--PNKIASGFFLRLLTS 811
Qy      506 KVLKNAKQACORLQGY-RMPF-AMAAATLFPDASGNLDKNARFSAIYQDSNKLSD-- 560
Db      812 KLMDIADI CKSLASFIKKPPDRGEVSMEDDTNKL-----MEVEDQSSNMLNNDVDP 865
Qy      561 -----DMUKLL-----ADPRKDEK 574
Db      866 SSVSDANEPGESQITGAINPLAEYLSKODLLFDMKLKFLCLCVTAQTNTVSPRADI 925
Qy      575 MAKLEVIILGNDDITIDNVSS--DFPNYVNSVIYPTQFETCSITPTTFE-VESFVVCIRPH 632
Db      926 RRLKMLIDS--STLEPTKSLHLMYLLKELGEBEY-----PLPMEDVLELLRPLSNV 977
Qy      633 TQPY-----TITNHLVYYPK-----YLKYSQKSF 658
Db      978 CSLYRRDQVCKTILNHVLAHVAKLQGSNMDSENTRDAGQGLTYIGAFMHLTKERKIYF 1037
Qy      659 AKARNIATCIB-FKQSDSDSQPLKCIYR--PGSPVETRSFAFVLLHH----- 705
Db      1038 SVRMALVNCVKTILLEADYSKALINVMGKQFVNVETQ--FLADNHQVRLAASIN 1095
Qy      706 -----ONPEFYDEIKLELPQ 721
Db      1096 RLFDQTKDSDSRLKALPLKLOQTAFFENAYLKAQGBREMSHSENPELTDEI----- 1148

```


Db 470 SNLESSOKSDILKLM-----NKIWCTIFRGISSQIOAENNGILGALIOGSLVEVDEPFMK 525
 Qy 304 L-BGSG-----SGLDYLPBLASABEABIKLKSEBRV- 336
 Db 526 LFTGASCRBSPAVCCLTLATTSIVGAAYKMGIEONMCEVNS-----FSLK-ESIMXK 579
 Qy 337 -LFTY-LDPPAQKLDFFSSABEV--KSFEEKFGRIILVKCNDLSFENLOCCA----- 383
 Db 580 LFPYQLEGD---LENSTEVPRIHSNFPRLVLEKILV-----SLTKMCKAAMNFPQSV 631
 Qy 384 -----ENEGETTVNEPEFV--TSLPDI-----KYNRKISADPHVDLHFSV 424
 Db 632 ECEHOKDEKELSPSEVEELFLOTFTDKMDFLTIVRECGIEKQSSIGFSVHQLKESID 691
 Qy 425 ROMIATTSPALMGSGEPTOS-----ALRGILHE-----A 454
 Db 692 RCLGLSBOULNNYSEBITNSETLVRC SRLVGLVLCYCYMGVIAEEAYKSELFOKANS 751
 Qy 455 AMQYPRQGI FSVYTCPPHPIFLVARIEKVLQSGITTHCAEPMKSSDSKYA-----Q 505
 Db 752 LMQCAESITLFPKKNBEBRIGSLANMO--LCTRLSCTKTS--FNKIASGFLLRLTS 809
 Qy 506 KYLKNAKQAQORIGQY-RMPF-AMAAATL FKQASGNLDKNARFSAIYRODSNKLSD-- 560
 Db 810 KLMNDIADICKSLASFYKPKFDEGEVSEMEDTNGNL-----MEVEDQSSMNLFPDYD 863
 Qy 561 -----DMKLTL-----ADFKPKR 574
 Db 864 SSVSDANBEGESOSTGAINPLAEYLSKODLLFLDMLFKLCVTTAQTNTVSFRAADI 923
 Qy 575 MAKLPYLGNLDTIDNVSSDFPNYVNSYIPTKQETCSKTPITE-----VEEF 625
 Db 924 RRLTLMLID-----STLEPTKSLHLMYLMLELGEBEYPLMEBV 966
 Qy 626 VPCIPRHTOPYTY-----TNHLYVPRYL--KYDSQS----- 657
 Db 967 LELKLPISVNCSLYRRDQVCKTILNHLVHLVNLQSNMDSBENTRDAOQFLTVIGAFM 1026
 Qy 658 -----FAKARNIAICIE-FKQSDBEDSOPLKCIYGR--PGGVFTRSAFAALH 705
 Db 1027 HLTERRKYIFSVMALVNLKTLLEADPYSKAMILNVMGDFVNEVFTQ--FLANHHQ 1084
 Qy 706 -----ONPEFY 711
 Db 1085 VMMLAESINRLFQDTRKSSRLKALPLKQOTAFENAYIKQEGEMREMSHAEPEL 1144
 Qy 712 DEIKLELPQLHEKHLHLTFPHV--SCDNSSKSGTKKQDVETQVGYMLPLKQGRV 769
 Db 1145 DEI-----YNRKSVLLTLIAVLVSCS-----PICERKQALPALCKSVKENG 1185
 Qy 770 TSEHPIPVSA-NLPSGLVGOELG--MGRHYG--PEIKWYDGGFRPLKISTH-LVSVVYQ 824
 Db 1186 --EPHLVKKVLEVYSETFGRRLLEDPMASHLDYLVLEMNLQDTEYNLSSFPILLANT 1242
 Qy 825 DOHLNFPQYOCOT-----ESGAQLGNELVKYLKSLHA--MEGVMIALFPTI 873
 Db 1243 --NIBFYRSCYVLLPHLVIRSHFDEVSIANOIOEDMKSLTLDFCPKILVNIILPYFAY 1300
 Qy 874 QLEFRVLTATQEEVAV-----NVTREYIIVHVAOCHIEGLES-- 909
 Db 1301 EGRDGMGAQORETATKVVDMLEKSENLGKQIDHLFISMLPEIIVELMTLHEPANSAS 1360
 Qy 910 -----HLBSYVKAAYKAEPYVASSEKTYHEELTKSMITLKRSAP 950
 Db 1361 OSTDLCDPSGDLDPADNPFPFBSHVIKATFA--YISNCHKT--KLSKILELSSPD- 1413
 Qy 951 LTNKLLKYSWFFDVLKISMAOHLIENSKVLLRQRFPAHYAHAVETVMMMLPHIQ 1010
 Db 1414 -----SYQXILAILCEQALETNNV--YKGRILIKIYH---LFSILLADI-- 1453
 Qy 1011 KPRDNEASKNANSHLAVFLIKRCFTFMDRGFVFKQIINNYISCFAPDPTLFEYKEEFLR 1070

Db 1454 -----KSGLGAMAFVLADVIYT-----LIHIYNORPSCIMDVIRS--FSLCCDLLS 1499
 Qy 1071 VYCNHEHIYPLNLPMPKRGRIQRQYODLODYSLDFECRHHFLVGLLREBGTALQER 1130
 Db 1500 QVC-----QTATYCKDALQNH-----HVIYGLI---PLVYQV 1532
 Qy 1131 EVRLIAISVLKULLIKHSFDDRYASRSHQARIATLYLPLFGLILEMFORINVRVDSPPPV 1190
 Db 1533 EYQXVLDLKLKVLIDNKDNE-----NLY-----ITIKLDDPP- 1566
 Qy 1191 NAGMTVQBSLAPVNPVLVTPQKSTLDSNLHKLQALASIGSIAPYTTSPNINSVARNA 1250
 Db 1567 -----DHVVFQDL-----RITQOKIX 1582
 Qy 1251 DSRGSLISTDGSNSLPERNSEKSNLDGKQOSTSGNSVNCBCKDOSEIKSL--LMQFL 1308
 Db 1583 YSRGPF-----SLLEBINHFLSVSV-----YDALPLTRLBGLKDLRQL 1621
 Qy 1309 YILKMSDDALFTYNNKASTSELMDFITISEVCLHQFOYMGKRYIARTG-----MMHAR 1362
 Db 1622 ELHKQWMDIM-----RASQDNPQDGIWVKLV--NLQLSGMAINHTEKEVLEAVGSC 1674
 Qy 1363 LOQGLSLD--NSLTENHSYGHSDADVYHQSLEBANIA--TEVCLTAL-DTL-----SL 1410
 Db 1675 LGEVGPIDFSTIAIOHSDKDASTKPL--KLFEDEKELQWTFIMLTIVANNLTVEDCVKVRSA 1732
 Qy 1411 PTLAFKNOLADHGNH--PLMKXVPDYVLCFLQKHQSTTA-----LKNVPTALRSL 1459
 Db 1733 AVTCKNLTATYTGHSFWEIYKMTTDBMLAYLOPRTIRKKEFLVPRFDKENEPEGLDI 1792
 Qy 1460 -----IYKPFSTYEGRADMCALCYEILK--CNSKLSIRTEASOLLYPLM 1505
 Db 1793 NLMIPLSENHDIWITLPCAFLDGSGTKCEIL--QLKPMCEVKTDFOQTVLPYLIHOL 1850
 Qy 1506 RNNEDYTKKSFVRTHLOVILISVOLADVVGIGSTRFOQSIINNCANSDRLIKHSF 1565
 Db 1851 LQDTNESW--RMLSTHVOGFPT-----SCLRHSQTSRSTTPNLT 1889
 Qy 1566 SSDVD-----LTKRIRTVLATAQMKHENDPEMLVDQYSLASVSTPELRKATWLD- 1619
 Db 1890 DSESHFRCCLDKKQSQRTMLAVVDYMRQRKPS-----SGTIFNDAPWIDL 1936
 Qy 1620 ---SMARIHVKNGLDSEAMCYVHTALV--AEYLTRKEAVOWEPRLPHSHSACLRSRG 1675
 Db 1937 NYLEVAKV-----AQSCAHTFALLVABYADKKSMDQD-----EKRS 1974
 Qy 1676 GYFRQCGTAFRYITPIDEASNMEDVGMQDVHFMEDVLMEL-----LEQCADGLM 1726
 Db 1975 LAFEBGQSOTTI--SLSLEKSKSEETGISLO-----DILLEIYRSGIBPDSLYCGGG-- 2024
 Qy 1727 KAERYELADYIKLIIPYERKRDPERLAHYDTLHAYASKYTEVMSGRRL----- 1778
 Db 2025 -----KMLQPI--TRLRTEYHBA--MMGRALVYTDLETALPSSTRQAGITIALON 2070
 Qy 1779 LETYRVAFFQGAQOYFTDSEIYVE-----GFPEDEGKEVIYKBP 1821
 Db 2071 LGLCHILSVYLGKDYEKNOCPELELHYQAAMRNQMDCTSVSKVEGTS--YHSBL 2128
 Qy 1822 LTPLEISORLLIKLYSDKFGSEBNVAMIDSGKVPKXIDSKYAYI----- 1866
 Db 2129 YNALQSLDRDEPSTYEBLSKYARVKEVEEMCK--RSLESYVSLYPTLSRLQALIELESI 2185
 Qy 1867 -----OYTHVAPFDEKELOERKTEFRSHNIR--FVPEMPFTQGGKQGVGEECK 1917
 Db 2186 GELFSRSVTH-----RQLSBVYIKWQXHSOLKDSQSFQEBPMAL- 2226
 Qy 1918 RTTILTAIHCFYVVKRIPWYOHHTDNLPIEVAIDEMSKYVABEFLQSCSSABVDMIR-- 1975
 Db 2227 -RTV-----LEIMEKEMDNSQREICIDITLTKHVELSILARFKNTQLPER 2273
 Qy 1976 --LOKLGQSVSVOVNAPPLAYARAF-----LDDTNTKRYDNRK 2014
 Db 2274 AIFOIKQYNSVSCGVSEWQLEBAQVFAKKEQSIALSTIKOMIKKLIDASCANMPS----- 2329

QY 2015 LKKEVRFQVEACGALA-----VNERLIKEDOLEYOENKANY-REMAKEL 2060
 DB 2330 -LKUTYTECLVRGCMWLAETCLENPAVINQTYLEKAVEVAVANVDESSDEL 2379

RESULT 12

US-08-874-266-2
 ; Sequence 2, Application US/08874266
 ; Patent No. 5955279

GENERAL INFORMATION:

APPLICANT: Gatti, Richard A.
 TITLE OF INVENTION: ATAXIA-TELANGIECTASIA: MUTATIONS IN THE ATM GENE
 NUMBER OF SEQUENCES: 33
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Knobbe, Martens, Olson and Bear
 STREET: 620 Newport Center Drive 16th Floor
 CITY: Newport Beach
 STATE: CA
 COUNTRY: USA
 ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FASTSEQ for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/874,266
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 ATTORNEY/AGENT INFORMATION:

NAME: Mays Veneko, Nancy

REGISTRATION NUMBER: 36,298
 REFERENCE/DOCKET NUMBER: UCLA006.006A
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619-235-8550
 TELEFAX: 619-235-0176
 TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 3056 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Protein
 FRAGMENT TYPE: Internal
 US-08-874-266-2

Query Match 1.6%; Score 172; DB 2; Length 3056;

Best Local Similarity 17.3%; Pred. No. 0.00019;
 Matches 435; Conservative 337; Mismatches 831; Indels 918; Gaps 118;

QY 66 LNDCLREMLLPYDFQTAIRROGRYICSTVPAKAEBAQSLFVTECIKTYS-DMHLV 124
 DB 271 LNDLSKVIYI---ELFQLOI-----YIHPRGAKTQHGK-----AYESTMRKI 311
 QY 125 NYKVED-----YSGEPRQLPNKVVKDKLPVHYVEDEVDKXEDASLGS 170
 DB 312 LNVLDLVNEISHISRGKYSGGFNNIAVKNLEIADICHOV-----FNEDETRSLBI 366
 QY 171 QKSGITKIGMVLKGNNSAISTVMSFKRFFHLIQLJDSGYKFEPLKD-LQKEPKGSIF 229
 DB 367 SCSYTTQ-----RSSDYVPCRRK-----TELJ-----WEYIKOHLQKSONDFDL 409
 QY 230 LGSFL-----YGVFRRNKV-----RRFAPELK-----MODKS 256
 DB 410 VPMLOIATOLISIKYRPSLNCSELPMLMISQLLPQRRGERPVRVLRCLTEVALCQDKR 469
 QY 257 SYLLAADSEVMEWTITINKLIQLNFEAMOKRNGDS-----HEDDEQSK 303

DB 470 SNLESSQKSDLLKLM-----NKIWCITFRGISSEIOIAENFGLLGAIQGSILVEVDREBWK 525
 QY 304 L-EGSG-----SGLDVLPBLAKSABABIKLKSRRYK- 336
 DB 526 LFTGSACRPSGCAVCCLTALTTSTIVPGAVKMGIEQNMCEVNR-----PSLK-BSIMKW 579
 QY 337 -LFF-LDPAOKLDPSAEPV--KSFEKRGKRLIVKCNDSLSPNLQCCVA----- 383
 DB 580 LFFQLEGD--LENSTVPPILHNPFLHVLKILV-----SLWKCKCKAMNPFQSV 631
 QY 384 -----ENEGPPTNVPPFV--TSLPDI-----KRNKISADPFDVLDNHSV 424
 DB 632 ECHNKKKEBELSPFBEVELFQTFDKNDFLIVRECGIEHGQSSISPSVQNLKESLD 691
 QY 425 ROMIATSPALMNGSGPFOS-----ALNGILHE-----A 454
 DB 632 RCLLSEOLANNYSSEITNSETLVRCRLVGVIGCYCMGVIAEEBAYVSELPQKANS 751
 QY 455 AMQYPRKQIFSVTCRHPDIFVARIEKVLQGSITHCAPYKMSDPSKVA-----Q 505
 DB 752 LMQAGESITLFPKNTNEBFRIGSLRMMQ-LCTRCLSNCTKKS-PNKIASGFLRLILTS 809
 QY 506 KVLKAKAQACQRLQY-RMPF-AMAPATLFDASGNLDKNAFSAITRQDSNKLSD--- 560
 DB 810 KLMDIADICKSLASFIPKPPDGRGESEMEDTNGNL-----MEVBDQSSMNLFNDDPD 863
 QY 561 -----DMKLK-----ADFRKREK 574
 DB 864 SSVSDANERGESQSTGAINPLAEYLSKODLFLDMLKFLCLCTTIAQTNTVSFRADI 923
 QY 575 MAKLPVILGNLDTIDNVSDFPNYSYIPTKOFETCSKTPYFE-----VEEF 625
 DB 924 RKKLMLLD-----SSTLEPTSLHNMVLMLEKLPGEERYLPMDV 966
 QY 626 VPCIPKTHPTIY-----TNHLYYPRKYL--KYDSQS----- 657
 DB 967 LELLPLSNVSLYRDQVCKTILNHVLAHVKNLQGSNMSENFTRDAQOGLTVIGAFW 1026
 QY 658 -----PAKARNAICIE-FKSDSEDSQPLKCIYGR--PGSFVTRSAFAVLHH 705
 DB 1027 HLTKEKRTIFSVRNALVNLKTLLEADPYSKWAILNVGKOPVNEVFTQ--FLDNHQQ 1084
 QY 706 -----ONPEFY 711
 DB 1085 VRLAASINLRFQDTKGDSSRLKALPLKIQDTAFENAYLKAQGRMESHASANPELT 1144
 QY 712 DEIKLEPTQLHEKHHLITFFPHV--SCDSSKSGTKRDVVEVQVGYSMPLDKGRNV 769
 DB 1145 DEI-----YNRKSVLTLIAVLVSCS-----PICKQALPALCKSVKENG- 1185
 QY 770 TSECHIPVSA-NLPSGYIGYQELG--MGRHYG-PEIKYVDGKFLPKLSTH-LVSTVYTG 824
 DB 1186 --EPHLVKVLEKVSFETGYRRLLEDPMASHLDYLVLEVLNLDQTEYNLSSPFIILANT- 1242
 QY 825 DOHLNPFQYCOKT-----ESGAQALGNELVLYKLSLHA-MEGHVIAIAPFTILN 873
 DB 1243 --NIDFPRSCYKVLPHLVIRSHFDEVKSIANOQGEOMKSLTLTCCPKILVNLIPRYAY 1300
 QY 874 QLFVLTATQEEVAV-----NTRVYIIVVAQCHEGLES-- 909
 DB 1301 EGTNRSGMAOORETATKYDMLKSENLGKQIDHLEFISNLEBEIVELMTLHEPANSSAS 1360
 QY 910 -----HLRSYVYKVKRPPVASYKYVHBLTSMNTILKPSADF 950
 DB 1361 QSTDLCDFSGDLDPANPNHPSPSHVTKYTPA--YISNCHK-----KLMSILBILSKSPD- 1413
 QY 951 LTSNKLKYSWFFEDVLIKSNAOHLIENSKYKLLRNQFPSPASYHAAVETVVMMLPHITQ 1010
 DB 1414 -----SYQKILAIICEQALETNNV--YKGRHILAIYH-----LPVSLLKDI-- 1453
 QY 1011 KFRDNPEASKANHSLAVIFRCKTFMDRGVFPKQINNYISCFAPGDEKTLPEYKFEFLR 1070
 DB 1454 -----KSLGAGAMFVLRDVIYT-----LIHYINRPSICINDVLSRS-FSLCDDLSS 1499

QY 1071 VCNHEHYIPLNPMFEGKRIQRYODLDYLTDEFCRNHEVLLEVTALQEFER 1130
1500 QVC-----QTAAYVCKDALENHL-----HYVGTLL-----PLVYEQV 1532
QY 1131 EVRLIAISVUKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVORINRVDSPPV 1190
1533 EVQKQVLDLKYLVINKDNE-----NLV-----ITIKLDPFP- 1566
QY 1191 NAGMTWKDESIALPAPNPLVTPQKSTLONSJLHKDLGALISGIASPYTTSTPINSVRA 1250
1567 -----DHVFKD-----RTQOKIK 1582
QY 1251 DSRGSLISTDSGNSLPERNSEKNSLDKIQOSSTLGNVVRCDKLDQSEIKSL-IMCPL 1308
1583 YSGGPF-----SLLEINHPLSVS-----YALPLTLLEGKDLRRL 1621
QY 1309 YLKSMSDALFTYMNKASTSEIMDFTTSEVCLHQFYMKRYIARTG-----MHAR 1362
1622 ELHKQDMVDIM-----RASQDNPDQDIMVLYV--NLQLSKMAINHTEKEVLEAVGSC 1674
QY 1363 LQQLGSLD--NSLTFNHSYGHSDADVHOSILEANIA--TEVCLTAL-DTL-----SL 1410
1675 LGEVGPIDSTIAIQSKDASATYKAL--KLPEDEKELQMTFIMLTLYLNTLVEDCVKVRSA 1732
QY 1411 FTLAFKQNLADHGHN--PLMKKVFVUYLCFLQKHQSETA-----LKNVFTALRSL 1459
1733 AVTCLKNILATKTHGHSFWEIYKWTTPMLAYLOPFRSTKRLFEVFRPKENFEBLDDI 1792
QY 1460 -----IYRPSTFEGRADMCALCYELIK--CNSKSSINTESAQLIYFLM 1505
1793 NLMIPLSENHDIWIKTLCAFLDSGGTKCEIL--QLKPMCEVKTPFCOTVLYPLIHDLL 1850
QY 1506 RNNFDYTGKKSFPRTLQVLIYSQILADVIGIGRRPOOSLIINNCAINSORLIGHSF 1565
1851 LQDTNSW--RNLSTHVOGFT-----SCLHFSQTSITTPPAVL 1889
QY 1566 SSDVKD-----LTKRITVLMATAQMKHEHNDPEMLVDQYSLAKSYASTPELARKTMD- 1619
1890 DSESEHFPCCLDKQSRMTLAVDYMRRQKRS-----SGTIFNDAMWLL 1936
QY 1620 ---SMARIHVKNQDLSBAMCYVHTALV--AEYLTRKEAVONEPRLPHSHSACLRSSG 1675
1937 NYLEVAKV-----AOSCAHFTALLYAEIYADKXSMDDQ-----EKRS 1974
QY 1676 GVPRQCTAFRVTTPNIDEASWMEVDGMDVHFNDDVLMEL-----LEQCADGLM 1726
1975 LAFEBSSOSTT--SSLEKSKKEETGISIQ-----DLLEIYRSIGEPDLSYCGCG- 2024
QY 1727 KAERYELADIYKLIPIYEKRRDFERLALHYDTLHRAYSKYTEVHNSGRRL----- 1778
2025 -----KMLQPI--TRULTYEHA--MMCKALVTYDLERAISSSTRQAGIIQALON 2070
QY 1779 LGTFYVAFFGQAQOYFTDSETDVE-----GFPEDEGKEIYKEPK 1821
2071 LGICHLISVYLKGLDYENDKMCPELELHYOAMRMNQDHTCSVSEKEVGS--YHEBL 2128
QY 1822 LRPLEISORILKLYGDKFGSENVKMIODSGKNRPDLDSKYAYI----- 1866
2129 YNHLQSLRBRFSTFYESLKVARKEVECK--SLSLSVSLYVTLTSLRLQALIGLESBI 2185
QY 1867 -----QVTHVLPFDEKELOERKTEPERSHNIR--FMFEMPFTQTSKQGVAEQCK 1917
2186 GELFSRSVTH-----RQLSEVYIKMQHSQLKXSDSFOERIAL----- 2226
QY 1918 RTIILTAICFPYVKKRIPYMTQHTDLAPLEIADMSKVAELQOLCSAEVDMIK-- 1975
2227 -RTVI-----LEILMEKENDSORBECIKDILTKHLEVELSILARTFKNQOLPBR 2273
QY 1976 --LQLLQGSVSVQVNAAGPLAVARAF-----LDDNTRKRYPNKKV 2014
2274 AIFQIKQVNSVSCGVSEMOLEBOVWAKKEOSLALSIUKOMIKKLDASCAANNPS--- 2329

QY 2015 LKKEVROFVENCQALA-----VNERLIKEQOLEYQOEMKANY--RENAKEL 2060
Db 2330 -LKTLYTECLARVCGNWLABTCLLENPAVIMQTYLEKAVAVAGYDESSDEL 2379
RESULT 13
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lymn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3159
; LENGTH: 10182
; TYPE: PRN
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3159
Query Match 1.5%; Score 167; DB 4; Length 10182;
Best Local Similarity 16.5%; Pred. No. 0.0046;
Matches 374; Conservative 335; Mismatches 868; Indels 688; Gaps 91;
QY 16 FTALSKGTAELRQSVSEVVRGSLAKPKLE-----PUDYENVIVOKTQILNDCL 70
Db 2461 PSANIKNNSQITPIPKAGQ---GNTENTNPTVIOAPAHQHTLTINEIKEGQANTNDI 2517
QY 71 REMLLPYDDPOTAILRRQRYI-----CSTVPA-----KAEEPAQSLFTECIKT 116
Db 2518 NNAVQVPP--KNRVALIQGNALPTNLGSGTSHIPVIYVSDGSSEBA---TEYVRT 2569
QY 117 YNSDHLVNYKEDVSGEFPRLPNKVYLDLPRVHVEVDEVDDEDAASIG--SQKGI 175
2570 KYNKTELINAR-----RRIDEISKENTPSSIRNPFQDMNMAQOINTMADADQV 2621
QY 176 TNGHGLYGNMNSALS--VTMRSEKRRFFHLIQLGDSYKFEFLKDQEKSGSIFLGF 232
Db 2622 IGTERATQOVNSALSKYQAQONKINAKALLQNKADNSQLVRKQEQDQSIQPAASDQ 2681
QY 233 LYGVSPNNKVRRAPEFLKMDKSSYLLADEVEMEETITLKIQLN-----FEAM 287
2682 MTQDSTRVYKNNRQAABQALOHANSVINNGDATSQ-----QINDAKNTVEQAO 2729
QY 288 QEKRGDGHEDDEQCKLEGSSGSLDSYVPELAKSAREALIKLSRSRYKLFYLPDPAKL 347
2730 RQVYEAQSNLRADKSQLOSAVDTLNRDV--LTNDKRPASVVRYNDA--ISNRKEIDTAKA 2786
QY 348 DFSS--AEEPVKSFEEKFGKRLIVKCNDSLFFNLQCVAAENEGPTTNVEFFVTLSLF 403
2787 DASSTLRMTNSVEGVRALAK-----INTVQD----- 2814
QY 404 DIKYNRKISADPHVDLNFVSQKMATTSPLAMNGSGEPTOSALRGILHEAMQYPKGI 463
Db 2815 -----KYNQALALQPKENNELQAKKRLQDADVND----- 2845
QY 464 FSVTCPRHDPILVARIEKVLGSIHCAEPYMKSSDSKVAQKYNKAKQACQRLGORYM 523
2846 -----IPQOQMTQOTTINNVDKQREARALTSQRYVINDQDATTOETSEKS 2893
QY 524 PFAMARTLFPDASGNL--DKNARFSAIYRQDSNKLSDNDMLKILADPRKPEKMAKLPIV 581
Db 2894 KVEQMQML--TNAKSNLADKNELOTAYNKLIEVSTNG-----KKRASIQYETA 2943
QY 582 LQNLDTITDYNSSDFPNYVNSSYIPTQFETCSKTPITFEVBEFVPCIPKHTQPYTIYN 641

Db 2944 KARIQNDANKNEERILIGNDNPVQS-----VTOANKIKAIQEPCLTEALIMLON 2995
 Qy 642 HUYVYKYLKYSOKSPAKARNIAICIEFKDSDEDSOPLKCYGRGPFVFTSAFAV 701
 Db 2996 K-----ENNTLVNAKRL-----ENAVNDTDP----- 3019
 Qy 702 LHHHONPER--YDEIKIELPTQJLHEKHLILLTFHVSCDSSKSGTKRDVV----- 751
 Db 3020 --HGMTOETINNYNAKKEAONEI--OKANMIINNMGATAQADISEKSEKYEVLQALONAK 3076
 Qy 752 -----ETQYGSYSLPLKKGGRVVTSEQHIFVSANLPSJYIGYELGMRHGYEIKR 803
 Db 3077 NDLRADKRELQTAYNLT-----IONVNTNKKPSSTIONY----- 3110
 Qy 804 VDGKPELKISTHVLVSTVY--TODQHLNFEOYQCKTESGAQ-----ALGNELYLXKS 855
 Db 3111 -----KSARNINQNTAKNEAHNVLENTPVNAVEDAKRKINAIQREVTAKAI 3162
 Qy 856 LHAMEGHVMIAPLPTILNQLFR--VLTRATQEEVAVNTVRIHVVAOCHEEGLESMLR 912
 Db 3163 LQDKEDNSBLVRAKEKLDQAINQPSLNGMTQESINNVTTK-----RR 3205
 Qy 913 SYVKVAYKAEPVY-----ASEKYVHEELTKSM-----TTILKPSADPL- 952
 Db 3206 EAQNIASSADTIINNGDASIEQITENKIRVEEATNALNEAKOHLTADTTLKTEVRKLSR 3265
 Qy 953 ---SNKLLKYSWFFDVLKISMAQHLIE--NSKYKLLRNOEFPAHYHAETVYVNMU--MP 1006
 Db 3266 RQDITNKKRSSVAVNNTIHSLOSETQTEENRANTIIKI-----RSVEEVNNAHENV 3320
 Qy 1007 HITQKFRD-----NPEASKNANHSIAVPIKRCFTFMDRGFEKOINNY----- 1049
 Db 3321 QLNQRLDTITNLQPLANKESLKEARNRLESKINE--TVQDQMTQOSYEVYQAKIKAO 3378
 Qy 1050 -----ISCFAQDDP-----KTLFEYKEFLRVCHENKPIPLN-PMFPGKRIQRYD 1097
 Db 3379 NESSIAQTLINNGDASQDEQVTEIEKLNQKSELTSIMHLYNKKEPLTAKNOLOANDI 3438
 Qy 1098 LQDYSLTQDFCNHPLVGLLREVGTAQOEFEVRLIAISVKNLLIHSFDDRYASRS 1157
 Db 3439 QK---PSTP-----GMTQOSVQSYERKLOEAK--DKINSINNVLANN----- 3475
 Qy 1158 HQARIAITLYPLFGLLIENVORINVDVSPFPVNAQMTVKDESLAL--PAVNPVLPQK 1214
 Db 3476 -----PDVNAIRTKVETEQOINNELTQAKQOLTVQKQPLINAKTALQ 3518
 Qy 1215 GSTLQNSLKHDLGALSGIASPTT-----STPNINS-----VRANDSRGSLSTDS 1261
 Db 3519 --SLDN-----QPSITMTETATIONVNAKROKAEQVIONANK---IENA 3558
 Qy 1262 GNSLPERNSEKNSLDKQOOSITGN--SVNRCDKLDQSEIKSLMCFIYIKSMDDAL 1319
 Db 3559 QPSVQOVSEKSK--VEQALSELNNAKSAALRADKQELCAVYNQL----- 3601
 Qy 1320 PTYVNAKASTSELMDFETISEVCLHQFOYMGKRYIARTGMHARLQOLGSLDNLSTFPHSY 1379
 Db 3602 -----QPTDLNNKKKPAISITVYNQRYQOFSSELMSTKTN--- 3634
 Qy 1380 GHSADAVLHQSLEANIATEVCLTALDTLSLFTLAFKQNLADHGNPLMKKYFDVY--- 1436
 Db 3635 --TDRLIKEQNPVADV-----NNALNKVREVOQKL 3663
 Qy 1437 ---LCLQKHOSBTALKQVFTALRSLLYKPSFTFEGRADMCAALCYELIKONSLSLSI 1493
 Db 3664 NEARALLQNKEDNSALVRAKEQLQAVDQVPSF--EG-----MQOQTKDYNSSQQA 3714
 Qy 1494 RTEASQLLYFLMKNPDYTGKKSFWRTHQVLIISVQLIADVVGIGTGFQOOSLINN 1553
 Db 3715 QOEISK-----AQOVIDNGDAITQOISNAKTVVERALBANA 3752
 Qy 1554 ANSDRLIKITSPSSDVKDLTKRIRTYLMTAQKKEHENDEPMIVDQYSLASVASTPEL 1613
 Db 3753 KTGLRADKE--ELONAVNQLTONIDTSGKTPASIRKY--NEAKSRIQTOIDSAKNEANS--- 3807

Qy 1614 RTWLDMSARJHVKNGDLSSEAMCYVHTALVAEYLTRKEAVQWEPPLYPHSHSACLRSS 1673
 Db 3808 -----ILNDNPQVS---QVTLA---LNKIKAVQ---PHLDKAIAMLYKJE 3844
 Qy 1674 RGVFRQCTARVITPNIIDEBASMEVQ-----MODVFNEDVLMEI 1717
 Db 3845 NNALVQAKQOQLQVNVNDVPTQGTDTTANNYKSKKEADELEIQAOQIINNDATEQ 3904
 Qy 1718 L-----EQCADGMKAEVRELLIDYKLIIPYEKRRDPERLAHLVDTLHRAISKYE 1770
 Db 3905 ITNETNRVQALINAIANKN-----DLRADQSQLENAYNQL-----ON 3943
 Qy 1771 VMSGRRLIG--TYFVAFEGQAOYQPTDSETDVGFPEDEDEKYEIYKEPRLTPLSEIS 1829
 Db 3944 VDTNKKRASIQQYQAA--RQALIEQYNNASSEAHQILIENS-----PSVNEVA 3990
 Qy 1830 QRLKLYSDKFGSENVKMIQDSGVNPKDLSKVAVIQVTHVIPFEDEKE---LOERT 1885
 Db 3991 QALQK-----VEAVQ-----LKVNDAILHONKENSALVTAKN 4024
 Qy 1886 EFERSHNIRRMFEMEPFOTGKQGVSE--QCKRTITLALHCFYVKKRIPWYQHHTD 1944
 Db 4025 QLOQSVN-----DQPLT--TGWTQDSINNYEAKREBAQSAL-----RMAEAVINNGD 4069
 Qy 1945 LNPFEVADENSK---KYAEIRQLCSSABVPMIKQLQLQOSVSQVNAAGPLAVARAFID 2001
 Db 4070 ATAKOIS--DESKVQOALHLNDADQOULTADTTELQTAQV-----QIAR-----RG 4114
 Qy 2002 DTNFK--RYPDNKYLKEVFRQFVEACGOALAVNERLIKEDOLEYQEBKANYREMAKE 2059
 Db 4115 DTNKKPRPSINAYNAKQISLETQITSADONNAVAIQKIFRTQVQVNRALQOVN--QLNQ 4172
 Qy 2060 LSEIMHQICPLEETSVLPNSIHFNALISGPISTWGHGTSSS 2104
 Db 4173 LTELIN--QLOPLSNNDALKAAKLNLNKINOT---VQDGTMQOS 4213

RESULT 14
 US-09-308-375-2
 ; Sequence 2, Application US/09308375
 ; Patent No. 6300117
 ; GENERAL INFORMATION:
 ; APPLICANT: Genencor International, Inc.
 ; TITLE OF INVENTION: Proteases from Gram-Positive Organisms
 ; FILE REFERENCE: GC394-PCT
 ; CURRENT APPLICATION NUMBER: US/09/308,375
 ; CURRENT FILING DATE: 1999-05-14
 ; EARLIER APPLICATION NUMBER: EP9719636.4
 ; EARLIER FILING DATE: 1997-09-15
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FaetsEQ for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 2285
 ; TYPE: prt
 ; ORGANISM: Bacillus subtilis
 US-09-308-375-2

Query Match 1.5%; Score 163; DB 4; Length 2285;
 Best Local Similarity 17.0%; Pred. No. 0.00066;
 Matches 405; Conservative 359; Mismatches 830; Indels 792; Gaps 99;

Qy 24 GNAALRSQSVSVNRGSLVLAKKPI--EPIDYE--NVIYVKQTOQ---LNDCLREMLLP 77
 Db 351 GAISGLKEMVSOALEIDTLMTNIRRVNRP--DYKINELLOSIDIGTLLSNKITDILQMT 409
 Qy 78 -----YDFQTAIRROGRYICSTVPAKAEBEASLFTVECTIKYV-----SDMHL 123
 Db 410 GDFGRMGDEBSLSTLTQAQVQVSDLTPTDPTNLT--TAAMLNFIANDSISINDK 467
 Qy 124 VNYKEDYSGFERQLPNKVVL--DKLPVHYVEVDEVDKEDDAISGQKGIYTHGWL 182
 Db 468 LNEVDNNYAVTTLNANSIRAGSGTASTFGVELNDLIGYTTAISTRESGIV----- 521

Db 2223 LTNLSLIPNLKKT-----EIPSKTIASSGDKTINLTFTFHDMLIGG 2263

RESULT 15

US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022

GENERAL INFORMATION:

APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022e1 Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CU 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-8949
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-328-254-6

Query Match 1.5%; Score 162; DB 1; Length 2482;

Best Local Similarity 17.3%; Pred. No. 0.00094;

Matches 422; Conservative 340; Mismatches 830; Indels 852; Gaps 109;

QY 49 IEPFLYENIVQKQVQIINDCLREMLFFYDDQVTAIILRQGRYICSTPAKAEERAGSL 108
DB 45 VEIRNLNHLNVLDSKSYEVETQKLAEMLOQKAEFS-----DQKHQKIEEM 89
QY 109 FTECIKTYNDSMHLVNVYKEDYSGEPRLPNKVVLDLTL--PVHY-YEVDEVDKDEDA 165
DB 90 -----CLKTSQ-----LTQGVDELDHGLQLSNEIMKDCYQDIAHAYLSLDLKSQA 140
QY 166 ASLGQSGKITTHGMLYKGNMNSAISVTNRSFKRRFFHLIQGDSY--KFEFLKDLQK 222
DB 141 SLVTNE-----DHQRSILAFDQOPAMHSHFANIIIGQSGMPERSECRLEADQ 188
QY 223 EKGSTFLGLYGVSTRNNKVRFAFELKQDKSSVTLAADSEVENHEVITTLNKLQAN 282
DB 189 SPKNSAIL-----QNRVDSLFLSLSQKQ---MNSDLQKQCEELVQIKGEL-EEN 234
QY 283 PEAAMQKNGSDHEDQSKLEGSGGLDSYLPFL--AKSAREARIKLKSSESVGLFYL 340
DB 235 LKAKBGMHQSFLAETSRISKQEDQTSANQNVAAETLSLLENKEKELQILND----- 286
QY 341 DPAQKLDPSASPEYKSFBEKFKRIILYKCNDSLNLQCVAAENEGPTTINVEPFVTL 400

Db 287 -----KVETQAEIQE-----LKKSNHLI-----EDSLKEQLLSBTL 319
QY 401 SLFDIKYNRKISADPHVDLNFPSVROMATATSPALMNGSGPETQSAI----- 447
Db 320 SL-----EKKEMSIISLNKREIELTQE-----NCTLKEINASLQOEKNLQKSES 367
QY 448 -----RGLHEAAMQPKQGIFFSVTCPPHDIPLVARIEKVLQSGITTHCAPEYMKSSD 499
Db 368 FANYIDEREKISIELSDQKQE-----KILLQRCBET-----GNAYEDLSQ 409
QY 500 SSKVAQKVLKNAKQAC-----QRLQGYRMPFA----- 526
Db 410 KYKAAQE--KNSKCECLNCTSLCENRKNLEQLKEFAKHQEFITKLAFABERNQL 467
QY 527 -----MAARTLPKASGLNDKNA-----RPSAIYRQDSNKL-----SNDMLKL 565
Db 468 MLEETVQALRSEKTDQNNNSKSAAGLKEIMTLKEQNMKEVNDLQEBQILKV 527
QY 566 LADFRKPRKMAKL PV-----ILGNLDITIDVSSDPFYVNSSVIPTQF 610
Db 528 MKTHCEQNLSBPIRNSGKERESERNQCNFPQWDLVKEISLD--SY-NAQLV---QL 581
QY 611 ETC--SKTPIPEVSEFVPCIPKATQPY--TITVHLVYVRYKLYKDSQ-KSPAKARIA 665
Db 582 EAMLNKKEIKLQESKEKEKELQHELOTIRGDLSTSNL-----QDMQSOEISGLKQCE 633
QY 666 ICIEFK-----DSDEDSQPLKCIYRGGPGVFTTSAPAAVLHHNQDEPFYDEI--- 714
Db 634 IDAEKTYISGPHLESTQNDNAHLQD-----SLQTTNKLNEIKICELIQA 680
QY 715 -KIELPQLAKHNLTLTFVYVSCDSSKSTKRDVETQVGYSMPLKXGRVVTSEQ 773
Db 681 EKYELVETLN-----DSRSECTAIRKMAE-EVG-----KLAVKVLINDS 721
QY 774 ---HIVSANTLPGYLYQ-----ELACMGNHYBEIKYVDCGRPLKISTHYL 817
Db 722 GLHGLVEDIPGGEFGQPNQEPHVSAPLDESNSYHLLTSLDK-----EVQNH 772
QY 818 VST-----VYTDOH--LHNPFOYCOKTESGAQALGNELVTKLSLHMEGMYMIAFLPTI 871
Db 773 AELQKPLSLOSEHKLND--QHCMSK-----MSELQTYVDSLKA-ENVLSTNLNLF 824
QY 872 LNOFLRVLTQAEVAVNVTVRIIIVVAQCHEGLESLSRYKVAKABPYVAEYKT 931
Db 825 QGDIVKEM-----QGLE----- 837
QY 932 VHEBITKSMITTLKPSADFLTSNKLKYSMFPFDVLLISMAOHLIENSKVALLNQRPPA 991
Db 838 --EGLVPLSLSSCVPDSSSLSS--LGDSSFYRLLEQTCGMSLSLNSDEGAVSANQ---- 888
QY 992 SYHNAVETVV--NMIMPHITQKFRNPREASKNAHSLAVFIKRCFTFMDRGVFVQINNY 1049
Db 889 ---CSVDVFCSSLOEENLTKR--ETPSAPAKGVEELSLQE-----VYNQSLK 933
QY 1050 ISCFAPGPKTLFEKFFELRVVNCNHEHYIPLNIMPFGKRIQRYQD-----QLDYS 1103
Db 934 L-----BEKMSQGIKMKKE-----IQLBQLLSERBELD-- 964
QY 1104 LTDFCRNHFVGLLIRVGTALQDFREVRLLAISVLKNLLIKHSFDRVARSHQARIA 1163
Db 965 -----CLRKQYLSENEQWQOKLTSTVTL-----EMESKLAEEKQTOL 1002
QY 1164 TLVYLPGLGLIENVGRINVRDVSPPRVNAGTVVDESIALPVPVLPQK-----GST 1217
Db 1003 SLELEVARLQGLD-LSSRSLLGITEDALQGRNESCDIKENSTETTERTPKHQVHOI 1061
QY 1218 LDNSLHKDL-----LGAISGIASPYTSTPINSVYRNADSR----- 1253
Db 1062 CDKXQQDVLNLDIEKITETGALKPTGECSSGSGSPDTNPERPGEDKTOGSSSEGISLSLSPG 1121
QY 1254 -GSLISTD--SGN-----SLPERNSEKNS-----LDKIQGSSTLGNSVVRCD-KLDQ 1297
Db 1122 PNALVPMDFLGNQEDIHNLQRLVYKETSNNENRLLHVLIEDRDKVYSILNEMKELDSKLHL 1181

QY 1298 SEIKSLMGLFYI-----LKMSDMLFTYMNKASTSELMDFFTISEVCLHOPQMG 1349
 1182 QEVQMTKIEACIELEKIVGELKENS-----LSEKLEYS-----CDH--QELL 1225
 QY 1350 KRIARTGM-----MHA---RLOQLGSLNLTFFNHSYGH-----SDADVL 1387
 1226 QRVETSEGLNLDLMEHADRSSREDIG--DVAKVNDSWKERFLUVENELSRIRSEKASIE 1283
 QY 1388 HOSL-LEAN---LATE-----VCLTALDLSLFTLAFKQQLADHGNPL 1428
 1284 HENLYLEADLEVQTEKCLEKDNENKOKYICLE--BELSVYT--SERNOL--RGELDT 1337
 QY 1429 MKKVFDVYLCLFLOKHOSLAKNVPALNSLIYKPPSTFYEGRADMCALCYEILKCNNS 1488
 1338 MSK-----KTTALDQISEKMEKTEGELES--HQSECLHCIOVAEAEVKEKTE 1382
 QY 1489 KLSISITSEASQLLYFLMRNNFDYTGKKSFYRTHLOVYISVQLIADVIGIGTRFQOSLS 1548
 1383 LLOTLSSDVSELTK-----DKTHLOEKLOSLEKDSQALSLTKCELENQIA 1427
 QY 1549 IINNCAUSDRLIGH-----LMTAQMKHEHNDPEMLVD-----TSFSDVXDL 1572
 1428 QUNK--EKELLVKESLSLOARLSESDYEKLNYSKALEALVEKGERPALRLSSTOEIVHOL 1485
 QY 1573 TKRIRTV-----LMATQMKHEHNDPEMLVD-----LOYSL-----A 1604
 1486 RRGIEKLRVRIADEKKQHLIAKLERENDSLDKVLENLERLOMSEBENGELVILDA 1545
 QY 1605 KSYASTPELRKTYLDSMAR-----IHKNGDLSAAMCYVHYT 1642
 1546 ENSKAETELKTOIEEMARSLKVFELDLVTLRSEKENLTQIOEKOGOLSELDKLSSPK 1605
 QY 1643 ALVAEYLTREKAVOWEPRLPHSHSACLRSSRGVFRGCGTAFRVITPNIDEASWMDV 1702
 1606 SLLEE---KEQAEIQ-----IKESKTAVENLQNOJKE----- 1635
 QY 1703 GMDVHFNEDV-----LMELBQCADGLWKAE-----RYELIADYKLIPI 1744
 1636 -----LNEAVALCGDDEIMKATEQSLDPRIBEHQLRNLSIEKLABLEADEKKQLC-V 1688
 QY 1745 YEKRRPFERLAHLYDTLHRAYSKVTEVMSGRRLTGTPRVAFPGQAQOYFTDSETV 1804
 1689 LQOLKESEHHAADLKGRENLERELIARTNOE-----HAALAEANSKGEVETLAKIE 1742
 QY 1805 GFPEDEDEGKEY-----IYKEPKLTPLSEISORLTLKYSKGESENVKMIODSGKVNPKD 1858
 1743 GMTQSLRGLELDVVTIRSEKENLTNELOKEKISELEIINSFENILOKEKQVOMKE 1802
 QY 1859 LDSKYAVIOVTHVIPPEDEKELQERKTEBERSHINIRFMFEMPFTQTKRGQGVBEQC-- 1916
 1803 KSTAMEMLOTOL-----KELNERVAAL--HN-----DQACAKA 1834
 QY 1917 KRTITITAHCPFYV-----KRIPLYQHTD--LNPIDVAIMSKVAEL-- 1962
 1835 KEONLSSQVECLELEKRAQLLOGLDEAKNNYIYLQSSVNGLIOVEDGKOLEKDEKIEIR 1894
 QY 1963 -----ROLGSAEVDMLKQLKLOGSAYOVNAGPLAYARAFID 2001
 1895 LKNQIODOQOLVSKSQVEGEHQLMKEONLELNLTVLELOKIQV-----LOSKNASLQ 1948
 QY 2002 DT-----NTKRYYPDNKVKULKEVFRQVACGOALAVNERLIKEDOL-----E 2044
 1949 DTLLEVQSSYKULLENLELTQDKMSFVE-----KVNMTAKETELQREHMHMAQKTAE 2002
 QY 2045 YQEBKAMRYEMAKELSEITMHE-----QICPLEEKTSVLPNSL 2092
 2003 LOBELSGEKNRLAGELQLLEIKSSKQDLKELTLENSELKXSL 2046

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:15:27 ; Search time 49.364 Seconds
(without alignments)
4965.877 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936
Sequence: 1 MSQPLLPASAEIRKFRAL.....ISGRTPTWVGMTSSSSV 2107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 11641973 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications AA.*
2: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pdb.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pdb.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pdb.*
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pdb.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pdb.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pdb.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pdb.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pdb.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pdb.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pdb.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pdb.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pdb.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pdb.*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10936	100.0	2107	9	US-09-815-379-8
2	10001.5	91.5	2066	9	US-09-978-244A-6
3	9926.5	90.8	1980	9	US-09-736-968A-108
4	9926.5	90.8	1980	10	US-09-736-968A-94
5	9926.5	90.8	1980	10	US-09-736-960-91
6	9886.5	90.4	2040	9	US-09-978-244A-27
7	7976	72.9	1534	10	US-09-736-968A-10
8	7976	72.9	1534	10	US-09-736-969A-10
9	7976	72.9	1534	10	US-09-736-960-10
10	6784.5	62.0	1352	9	US-09-736-968A-9
11	6784.5	62.0	1352	10	US-09-736-969A-9
12	6784.5	62.0	1352	10	US-09-736-960-9
13	6375.5	58.3	2073	9	US-09-978-244A-10
14	6105.5	55.8	2008	9	US-09-736-968A-105
15	6105.5	55.8	2008	9	US-09-736-968A-105
16	6105.5	55.8	2008	10	US-09-736-969A-2
17	6105.5	55.8	2008	10	US-09-736-969A-2
18	6105.5	55.8	2008	10	US-09-736-969A-91
19	5651	51.7	2175	9	US-09-978-244A-4

20	5560	50.8	2180	9	US-09-736-968A-110	Sequence 110, App
21	5560	50.8	2180	9	US-09-978-244A-2	Sequence 2, Appl1
22	5560	50.8	2180	10	US-09-736-969A-96	Sequence 96, Appl
23	5560	50.8	2180	10	US-09-736-960-93	Sequence 93, Appl
24	3608.5	33.0	1214	9	US-09-736-968A-12	Sequence 12, Appl
25	3608.5	33.0	1214	9	US-09-736-969A-8	Sequence 8, Appl1
26	3608.5	33.0	1214	10	US-09-736-969A-12	Sequence 12, Appl
27	3608.5	33.0	1214	10	US-09-736-960-12	Sequence 12, Appl
28	3406.5	31.1	1288	9	US-09-736-968A-13	Sequence 13, Appl
29	3406.5	31.1	1288	10	US-09-736-969A-13	Sequence 13, Appl
30	3406.5	31.1	1288	9	US-09-736-960-13	Sequence 13, Appl
31	2888	26.4	738	9	US-09-736-968A-11	Sequence 11, Appl
32	2888	26.4	738	10	US-09-736-969A-11	Sequence 11, Appl
33	2888	26.4	738	10	US-09-736-960-11	Sequence 11, Appl
34	2460	22.5	2045	9	US-09-736-968A-109	Sequence 109, App
35	2460	22.5	2045	10	US-09-736-969A-95	Sequence 95, Appl
36	2460	22.5	2045	10	US-09-736-960-92	Sequence 92, Appl
37	2460	22.5	2045	9	US-09-736-968A-2	Sequence 2, Appl1
38	2360	21.6	2047	9	US-09-978-244A-31	Sequence 31, Appl
39	2345	21.4	2090	9	US-09-736-968A-107	Sequence 107, App
40	2345	21.4	2090	9	US-09-978-244A-28	Sequence 28, Appl
41	2345	21.4	2090	10	US-09-736-969A-93	Sequence 93, Appl
42	2345	21.4	2090	10	US-09-736-960-90	Sequence 90, Appl
43	2325	21.3	1999	9	US-09-978-244A-8	Sequence 8, Appl1
44	2242	20.5	2032	9	US-09-978-244A-12	Sequence 12, Appl
45	2237.5	20.5	2015	9	US-09-736-968A-106	Sequence 106, App

ALIGNMENTS

RESULT 1
US-09-815-379-8
; Sequence 8, Application US/09815379
; Publication No. US20030073613A1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; TITLE OF INVENTION: GERITEN, MARY
; TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: 10716/35
; CURRENT APPLICATION NUMBER: US/09/815.379
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/191,134
; PRIOR FILING DATE: 2000-03-22
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2107
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-379-8

Query Match 100.0%; Score 10936; DB 9; Length 2107;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSQPLLPASAEIRKFRALSKPGTAAELRQSVSVVAGSVTLARPKLIBPDYENVYQ 60
DB 1 MSQPLLPASAEIRKFRALSKPGTAAELRQSVSVVAGSVTLARPKLIBPDYENVYQ 60
QY 61 KKTQILNCLCEMLFPDDPOTATLRQGRYICSTVPAKKEEQAQSLPTEICITVNSD 120
DB 61 KKTQILNCLCEMLFPDDPOTATLRQGRYICSTVPAKKEEQAQSLPTEICITVNSD 120
QY 121 WHLVNYKYEDYSGEERQLPNKVVKDKLPVHYVEDEVDXDBAASLGSQKGIITKGM 180
DB 121 WHLVNYKYEDYSGEERQLPNKVVKDKLPVHYVEDEVDXDBAASLGSQKGIITKGM 180
QY 121 WHLVNYKYEDYSGEERQLPNKVVKDKLPVHYVEDEVDXDBAASLGSQKGIITKGM 180
DB 121 WHLVNYKYEDYSGEERQLPNKVVKDKLPVHYVEDEVDXDBAASLGSQKGIITKGM 180
QY 161 LYKGMNAISVYMASFGRRFFHLIQLDGSKYKPEFLDKQKPKGSIPLGFLYGVSPRN 240
DB 161 LYKGMNAISVYMASFGRRFFHLIQLDGSKYKPEFLDKQKPKGSIPLGFLYGVSPRN 240

241 NKVRFAFELKQDKSSYLLAADSEVEMEWITLTKIQLNEEAMOEKRNDSHEDD 300
 241 NKVRFAFELKQDKSSYLLAADSEVEMEWITLTKIQLNEEAMOEKRNDSHEDD 300
 301 OSKLBGSGGSDSYLPELAKSAREAEIKLKSESRYVLPLYDPAQCLDPSSAPPEVKSFE 360
 301 OSKLBGSGGSDSYLPELAKSAREAEIKLKSESRYVLPLYDPAQCLDPSSAPPEVKSFE 360
 361 BKRGKRLVKCNDLSFNLOCCVANESEPTTNEPFFVLLSLPDIKYNKISADPFVLDN 420
 361 BKRGKRLVKCNDLSFNLOCCVANESEPTTNEPFFVLLSLPDIKYNKISADPFVLDN 420
 421 HESVRQMIATTSBALNNGSGPETQSALRGILHEAAMQPKGIFSYTCBPDIFLVARI 480
 421 HESVRQMIATTSBALNNGSGPETQSALRGILHEAAMQPKGIFSYTCBPDIFLVARI 480
 481 KVLQSGITTCABEYKMSDSSKVAQVYLKNAKQACRLQCYRMPFMAARTLFKDASGUL 540
 481 KVLQSGITTCABEYKMSDSSKVAQVYLKNAKQACRLQCYRMPFMAARTLFKDASGUL 540
 541 DKNAFSAIYRQDSNKLSDNDMLKCLADFRKPEKMAKLPVILGNLDTITDYNSSDPENY 600
 541 DKNAFSAIYRQDSNKLSDNDMLKCLADFRKPEKMAKLPVILGNLDTITDYNSSDPENY 600
 601 NSSYIPFKQFETCSKPTITEVESEFVPCIPKHTQPYTIYTNHLVYVPKYLKXDSQKSPAK 660
 601 NSSYIPFKQFETCSKPTITEVESEFVPCIPKHTQPYTIYTNHLVYVPKYLKXDSQKSPAK 660
 661 ANRIACIEFKOSDEEDSOPKLCTYRPGQPVFTSAPAAVLHHONPEFYDEIKELPT 720
 661 ANRIACIEFKOSDEEDSOPKLCTYRPGQPVFTSAPAAVLHHONPEFYDEIKELPT 720
 721 OLHEKHLTLTPFVHSCDSSKSGSTKRRDVEYQVYSMLPLKDDRVTSSEQHTVSN 780
 721 OLHEKHLTLTPFVHSCDSSKSGSTKRRDVEYQVYSMLPLKDDRVTSSEQHTVSN 780
 781 LPSGVLGYOELGNGRYGPEIKKWDGKPLKISTHVLSTVYVTOQHLNFPQYCOKTES 840
 781 LPSGVLGYOELGNGRYGPEIKKWDGKPLKISTHVLSTVYVTOQHLNFPQYCOKTES 840
 841 GAOALCNELVKYIKSLHAMEGHVIAFLPTILNQLFRVLTRATOEBAVANTRVYIHHVA 900
 841 GAOALCNELVKYIKSLHAMEGHVIAFLPTILNQLFRVLTRATOEBAVANTRVYIHHVA 900
 901 OCHBEGLESHTLSYVYAKABEYVASEKTYHEELTKSMTTILKXSADPLFNSKLLKYS 960
 901 OCHBEGLESHTLSYVYAKABEYVASEKTYHEELTKSMTTILKXSADPLFNSKLLKYS 960
 961 WFFPDVLIKSMQHLIENSKVLLRNQRPASVYHAAVETVVMMLMPEHITQKFRDNPEAK 1020
 961 WFFPDVLIKSMQHLIENSKVLLRNQRPASVYHAAVETVVMMLMPEHITQKFRDNPEAK 1020
 1021 NANHSLAVITKRCFTMDRGFVFKOINNYISCPADDPKTLFEYKEEPLRVACNHEHYIP 1080
 1021 NANHSLAVITKRCFTMDRGFVFKOINNYISCPADDPKTLFEYKEEPLRVACNHEHYIP 1080
 1081 LNLPMFPGGRIGRQYDLOLDVSLTDEFCRNHFLVGLLREVTALAOEFREYVLIASVL 1140
 1081 LNLPMFPGGRIGRQYDLOLDVSLTDEFCRNHFLVGLLREVTALAOEFREYVLIASVL 1140
 1141 KULLIHSFDDRYASRSHQARIATLPLFGLLIENVORINVDVSPFVNAGMTVXDS 1200
 1141 KULLIHSFDDRYASRSHQARIATLPLFGLLIENVORINVDVSPFVNAGMTVXDS 1200
 1201 LALPAVNPVLTPOKSGTLDNSLHKDLGAIISGASPYTTSTPNINSVRNADSGSLISTD 1260
 1201 LALPAVNPVLTPOKSGTLDNSLHKDLGAIISGASPYTTSTPNINSVRNADSGSLISTD 1260
 1261 SGNLSLBERSEKSNLSJDKQOQSTLGNVYRCDKLOSEIKSLMLKFLYILKSMSSDALP 1320
 1261 SGNLSLBERSEKSNLSJDKQOQSTLGNVYRCDKLOSEIKSLMLKFLYILKSMSSDALP 1320
 1321 TYWNKASTSELMDFPTISEVCLHQFOYMGKRYIARTGMHARLOQLGSLDNSLTFNHSYG 1380
 1321 TYWNKASTSELMDFPTISEVCLHQFOYMGKRYIARTGMHARLOQLGSLDNSLTFNHSYG 1380

1321 TYWNKASTSELMDFPTISEVCLHQFOYMGKRYIARTGMHARLOQLGSLDNSLTFNHSYG 1380
 1381 HSDADVHQSLLSEANIAIEVCLTALDITSLPTLARKNOLDHGNPMLKRYFDVYLCEL 1440
 1381 HSDADVHQSLLSEANIAIEVCLTALDITSLPTLARKNOLDHGNPMLKRYFDVYLCEL 1440
 1441 OXHOSETALKNVFTLRSILYKFPSTPYEGRADMCALCYELKCKNSKLSIRTEASOL 1500
 1441 OXHOSETALKNVFTLRSILYKFPSTPYEGRADMCALCYELKCKNSKLSIRTEASOL 1500
 1501 LYFLMRNNDYTGKSPVRTHLOVILISVQLIADVYIGGTRFQOSLSIINNCAUSDRI 1560
 1501 LYFLMRNNDYTGKSPVRTHLOVILISVQLIADVYIGGTRFQOSLSIINNCAUSDRI 1560
 1561 KATSSSDVYKDLTKIRIYVLAATAQMKHENDPEMLVLOYSLASVASTPLRTWTWDS 1620
 1561 KATSSSDVYKDLTKIRIYVLAATAQMKHENDPEMLVLOYSLASVASTPLRTWTWDS 1620
 1621 MARIHVQGDLSSEAMCYVHTALVAEYLTRKEAVQWEPPLPHSHSACLRSGVFRQ 1680
 1621 MARIHVQGDLSSEAMCYVHTALVAEYLTRKEAVQWEPPLPHSHSACLRSGVFRQ 1680
 1681 GCTAFRVITPNIDEBASNMEDVGMQDVHFNEDVLMELLBOCADGLMKAREYELIADYKL 1740
 1681 GCTAFRVITPNIDEBASNMEDVGMQDVHFNEDVLMELLBOCADGLMKAREYELIADYKL 1740
 1741 IIPYIEKRDPFRLAHLVDTLHRAVSKTYEVHSGRLLGTYFRVAFQQAAYOFTDSE 1800
 1741 IIPYIEKRDPFRLAHLVDTLHRAVSKTYEVHSGRLLGTYFRVAFQQAAYOFTDSE 1800
 1801 TDVEGFEDDEDEKXIYKEPKLTPLEISORLLKLYSDPKGSENVKMIODSKVNPOLD 1860
 1801 TDVEGFEDDEDEKXIYKEPKLTPLEISORLLKLYSDPKGSENVKMIODSKVNPOLD 1860
 1861 SKYAVIYQVTHVIPFDEKELDERKTEPERSHNIIRFMEEMPTQTKQOGVEOCKRRT 1920
 1861 SKYAVIYQVTHVIPFDEKELDERKTEPERSHNIIRFMEEMPTQTKQOGVEOCKRRT 1920
 1921 IITAIHCPYVYKRIIPWYQHTTDLNPIEVALIDENSKVYAEIROCSSAEVDMIKLOLKL 1980
 1921 IITAIHCPYVYKRIIPWYQHTTDLNPIEVALIDENSKVYAEIROCSSAEVDMIKLOLKL 1980
 1981 OGSVSQVWAGPLAARAFDLDTNTRKYPDNKVKLLKEVFRQVFAACGALAVNERLKE 2040
 1981 OGSVSQVWAGPLAARAFDLDTNTRKYPDNKVKLLKEVFRQVFAACGALAVNERLKE 2040
 2041 DQLEYOEKWKANYREMAKELSEIMHEQICPLEKTSVLPNSLHIFNAISGTFSTWVHGM 2100
 2041 DQLEYOEKWKANYREMAKELSEIMHEQICPLEKTSVLPNSLHIFNAISGTFSTWVHGM 2100
 2101 TSSSSSV 2107
 2101 TSSSSSV 2107

RESULT 2
 US-09-978-244A-6
 ; Sequence 6, Application US/09978244A
 ; Publication No. US20030103992A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lu, Peter S
 ; APPLICANT: Garman, Jonathan D.
 ; APPLICANT: Candia, IIT, Albert P.
 ; APPLICANT: Arbor Vita Corporation
 ; TITLE OF INVENTION: CLASP MEMBRANE PROTEINS
 ; FILE REFERENCE: 020554-000161US
 ; CURRENT APPLICATION NUMBER: US/09/978,244A
 ; CURRENT FILING DATE: 2001-10-15
 ; PRIOR APPLICATION NUMBER: US 60/310,028
 ; PRIOR FILING DATE: 2001-08-03
 ; PRIOR APPLICATION NUMBER: US 09/737,246
 ; PRIOR FILING DATE: 2000-12-13

PRIOR APPLICATION NUMBER: US 09/736,969
PRIOR FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 09/736,960
PRIOR FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 09/736,968
PRIOR FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 60/240,545
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,508
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,503
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,539
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,543
PRIOR FILING DATE: 2000-10-13
Remaining Prior Application data removed - See file wrapper or PALM.
NUMBER OF SEQ ID NOS: 106
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 6
LENGTH: 2066
TYPE: PRT
ORGANISM: Mus sp.
FEATURE:
OTHER INFORMATION: CLASP-2
US-09-978-244A-6

Query Match 91.5%; Score 10001.5; DB %; Length 2066;

Local Similarity 92.5%; Pred. No. 0;
Match 1930; Conservative 49; Mismatches 57; Indels 51; Gaps 8;

51 PUDENVIVQKKTQILNDCLREMLLPYDDPOTAILRQARYICSTVPAKAEESQSLFV 110
1 PUDENVIVQKKTQILNDCLREMLLPYDDPOTAILRQARYICSTVPAKAEESQSLFV 60
111 TECITVNSDMLVNYKYEDYSGEPRQLPNKVYKDKLPVHYEVDEVDKEDAAASLGS 170
61 TECITVNSDMLVNYKYEDYSGEPRQLPNKVYKDKLPVHYEVDEVDKEDAAASLGS 120
171 QKGGITTKGWLKYGNNNSAIVTMSFKRPFHLIQLGDSYNLNFYKDEKSKCKGSI 228
121 QKGGITTKGWLKYGNNNSAIVTMSFKRPFHLIQLGDSYNLNFYKDEKSKCKGSI 180
229 FLGFLYGVSRNNKYARFAPFLMODKSSYLLAADSEVEEWEITLTKILOLNFPAHQ 288
181 FLGFLYGVSRNNKYARFAPFLMODKSSYLLAADSEVEEWEITLTKILOLNFPAHQ 239
289 EKRNDSDHEDEQSKLEGGSGSLDSYLPFLAKSARBAIILKSEBSRVKLFYLDPAQKLD 348
240 EKRNDSDHEDEQSKLEGGSGSLDSYLPFLAKSARBAIILKSEBSRVKLFYLDPAQKLD 299
349 PSSASPEVKSFEKKGKRLIVKCNLDSFNLCCVANEGERPTTNNPFPVITSLPDIKN 408
300 PSSASPEVKSFEKKGKRLIVKCNLDSFNLCCVANEGERPTTNNPFPVITSLPDIKN 359
409 RKISADFHVDLNFVSROMIATTSAPLMNG--SGPETQGLRGILHEAAMOTPKOIFSV 466
360 RKISADFHVDLNFVSROMIATTSAPLMNG--SGPETQGLRGILHEAAMOTPKOIFSV 414
467 TCPHPDIFLVARIKYLQGSITHCAPYKSSSDSRYAOKVILKNKAOQORLGOYMPRA 526
415 TCPHPDIFLVARIKYLQGSITHCAPYKSSSDSRYAOKVILKNKAOQORLGOYMPRA 474
527 WAARTLFCOASGNLMDNARFSAIYRDSNKLSDMDLKLADPRKREKAKLPIVILGND 586
475 WAARTLFCOASGNLMDNARFSAIYRDSNKLSDMDLKLADPRKREKAKLPIVILGND 534
587 ITIDVNSDPPNYVNSYIPTKOFETCSKTPITFEVEFVPCIPKHTQPTTYITNMLYY 646
535 ITIDVNSDPPNYVNSYIPTKOFETCSKTPITFEVEFVPCIPKHTQPTTYITNMLYY 594
647 PLYLYKDSQKSPAKARNIAICIEFKDSDEDSQPLKCIYRPGGPVFTRSAPAAVLHHQ 706

595 PLYLYKDSQKSPAKARNIAICIEFKDSDEDSQPLKCIYRPGGPVFTRSAPAAVLHHQ 654
707 NPEFYDEIKIELPOTJHEKHMLLTFPHVSCNSKSGSTKRDVETQVGYMPLKXG 766
655 NPEFYDEIKIELPOTJHEKHMLLTFPHVSCNSKSGSTKRDVETQVGYMPLKXG 714
767 RVVTSEQHI PVSANLPSGYLGQELGMRHYGPEIKWYDGGKPLKISTHLVSTVYTDQ 826
715 RVVTSEQHI PVSANLPSGYLGQELGMRHYGPEIKWYDGGKPLKISTHLVSTVYTDQ 774
827 HLANFPOYCOKTESAQALGNELVYKLSLAMEGHVIAFLPTLNDLFRVLTATQOE 886
775 HLANFPOYCOKTESAQALGNELVYKLSLAMEGHVIAFLPTLNDLFRVLTATQOE 834
887 VAVNTRVYIHVVAOCHEGLESLSRYKAYKAPVYASRYKTVHELTSMITLKP 946
835 VAVNTRVYIHVVAOCHEGLESLSRYKAYKAPVYASRYKTVHELTSMITLKP 894
947 SADFLTSNKLKYSWFFDVLKSMOHLIENSKVKLNRORFPASVYHNAVETVNMIMP 1006
895 SADFLTSNKLKYSWFFDVLKSMOHLIENSKVKLNRORFPASVYHNAVETVNMIMP 954
1007 HITOKFRDNPENASKANSLAVFIKCTFPMDRGVFPQIINNYISCFAPGDKTLEBYKF 1066
955 HITOKFRDNPENASKANSLAVFIKCTFPMDRGVFPQIINNYISCFAPGDKTLEBYKF 1014
1067 EFLRVVCHNEHYIPLNMPFGKRIORYODQLDYSLTDECRNHPVLGILLREVTAL 1126
1015 EFLRVVCHNEHYIPLNMPFGKRIORYODQLDYSLTDECRNHPVLGILLREVTAL 1074
1127 QEFREVRILAIISVLNKLKISFDDRYASRSHOARIATLYLPGLIENVORINRVDS 1186
1075 QEFREVRILAIISVLNKLKISFDDRYASRSHOARIATLYLPGLIENVORINRVDS 1134
1187 PEPVAGMTVYDESLALPAVNPVTPQKSTIDNSLHODLCAIGIASPTTSPNINS 1246
1135 PEPVAGMTVYDESLALPAVNPVTPQKSTIDNSLHODLCAIGIASPTTSPNINS 1194
1247 VVNAADSRGLISTDSGNSLPERNSEKNSLDDHQOOSTLGSNVVNCDLKDOSEISLMLC 1306
1195 VVNAADSRGLISTDSGNSLPERNSEKNSLDDHQOOSTLGSNVVNCDLKDOSEISLMLC 1253
1307 FLYILKNSDDALFYNNKASTSELMDEFTISEVCLHFOYVNGKRYIA----- 1354
1254 FLYILKNSDDALFYNNKASTSELMDEFTISEVCLHFOYVNGKRYIA----- 1313
1355 -----RTGMMHARLQGLSDNSLTFNHSYGHSDADVHLQSLLEANIATEVCIT 1403
1314 RKSQTL PVSRRRTGMMHARLQGLSDNSLTFNHSYGHSDADVHLQSLLEANIATEVCIT 1373
1404 ALDTLSLFTLAFKNOGLADHGNPLMKVFPDYLCELGHOSETLAKVFTLRSGLIYKF 1463
1374 ALDTLSLFTLAFKNOGLADHGNPLMKVFPDYLCELGHOSETLAKVFTLRSGLIYKF 1433
1464 PSTFYEGRADMCAALCYEILKCNKNSKSSIRTEASQLLYFLMRNNPDYTGKSFVYTHLO 1523
1434 PSTFYEGRADMCAALCYEILKCNKNSKSSIRTEASQLLYFLMRNNPDYTGKSFVYTHLO 1493
1494 VIISVQILADVIGIGTRFOOSLSIINNANSDRILIKHTSPSSDVOLTKRIRTYLMAV 1553
1584 AOMKEHNDPEMLVLOYSLAKSYASTBELRTKMLDSMARIVHKNGDSLEAMCVYHVA 1643
1554 AOMKEHNDPEMLVLOYSLAKSYASTBELRTKMLDSMARIVHKNGDSLEAMCVYHVA 1613
1644 LVAEYLTKE--AVQMEPPLPHSHSACLRRSGVROGCTARVITPNIIDEASVME 1700
1614 LVAEYLTKEADLALQRRPAPFYSHSCTQKSKWGMROGCTARVITPNIIDEASVME 1673
1701 DVGMQDVHFNEDVLMELLEOCADGLWKAERYELLADYKLIIPYKGRDPFRLAHLYDT 1760
1674 DVGMQDVHFNEDVLMELLEOCADGLWKAERYELLADYKLIIPYKGRDPFRLAHLYDT 1733

QY 1761 LHRVSKTEVMSGRRLLGTYPFVAFGQAAQYQFTDSETEGFEDEGKEIYTKEP 1820
1734 LHRVSKTEVMSGRRLLGTYPFVAFGQ-----GFEDEGKEIYTKEP 1779
QY 1821 KLTPLSEISORLKLKYSDFKGSENVMIQDSGVNPKDLDSKAYIQTHTVIFPFDEKEL 1880
1780 KLTPLSEISORLKLKYSDFKGSENVMIQDSGVNPKDLDSKAYIQTHTVIFPFDEKEL 1839
QY 1881 QERKTEFSRSHNRIRMEFEPFQTQGRQGVGEQCKRTILAIHCFFYVKRIEVMQ 1940
1840 QERKTEFSRSHNRIRMEFEPFQTQGRQGVGEQCKRTILAIHCFFYVKRIEVMQ 1899
QY 1941 HHTDLNPIEVAIDEMSKVAELRQLQSSAEDMIKQLKQSSVSQVNAAPLAVARAF 2000
1900 HHTDLNPIEVAIDEMSKVAELRQLQSSAEDMIKQLKQSSVSQVNAAPLAVARAF 1959
QY 2001 DDNTKRYPDNKKLKEVFRQVEACGALLANNEKLKEDQLEYOEEMKANYRENAKEL 2060
1960 DDNTKRYPDNKKLKEVFRQVEACGALLANNEKLKEDQLEYOEEMKANYRENAKEL 2019
QY 2061 SEIMBOICPLEEKTSTVLPNSLHFNIAISGTPSTVWGHWTSSSVV 2107
2020 SDIMBOICPLEEKTSTVLPNSLHFNIAISGTPSTVWGHWTSSSVV 2066
Db

RESULT 3
US-09-736-968A-108
Sequence 108 Application US/09736968A
Patent No. US20020169283A1
GENERAL INFORMATION:
APPLICANT: Lu, Peter
APPLICANT: Garman, Jonathan David
APPLICANT: Candia, III, Albert Frederick
APPLICANT: Arbor Vita Corporation
TITLE OF INVENTION: CLASP-7 Transmembrane Protein
FILE REFERENCE: 020054-00061108
CURRENT APPLICATION NUMBER: US/09/736,968A
CURRENT FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 60/160,860
PRIOR FILING DATE: 1999-10-21
PRIOR APPLICATION NUMBER: US 60/162,498
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: US 60/170,453
PRIOR FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: US 60/176,195
PRIOR FILING DATE: 2000-01-14
PRIOR APPLICATION NUMBER: US 60/182,296
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: US 09/547,276
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,267
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,460
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,527
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,528
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 09/687,837
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,503
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,508
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,539
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,543
PRIOR FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 115
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 108
LENGTH: 1980

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human CLASP-2
US-09-736-968A-108
Query Match 90.8%; Score 9926.5; DB 9; Length 1980;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1932; Conservative 8; Mismatches 15; Indels 105; Gaps 6;
73 MLFPYDPEQTAIARQGRYICSTYPAKAEBAOSLFYECIKTNSMDHLVNYKEDXS 132
1 MLFPYDPEQTAIARQGRYICSTYPAKAEBAOSLFYECIKTNSMDHLVNYKEDXS 60
QY 133 GEFROLPNKVVLDLPHVYVEDEVDKEDBAALSGQKGIITGHGMLYKGNMSAISV 192
61 GEFROLPNKVVLDLPHVYVEDEVDKEDBAALSGQKGIITGHGMLYKGNMSAISV 120
QY 193 TWRSEFRREFFHILQIGDGSYKFEFLKD--LQKPKGISIFLGLYGVSPRNNVRRAPEL 250
121 TWRSEFRREFFHILQIGDGSYKFEFLKD--LQKPKGISIFLGLYGVSPRNNVRRAPEL 179
QY 251 KMQDKSSYLLADSEVMEEMWITINKILQNLFEAMQEKRGDSDHEDEQSKLGSGSG 310
180 KMQDKSSYLLADSEVMEEMWITINKILQNLFEAMQEKRGDSDHEDEQSKLGSGSG 239
QY 311 LDSYLPDLAKSAREAEIKLKSSESVKLFYLDPAQKLDSSAEPKVSFEKFRKRIILYK 370
240 LDSYLPDLAKSAREAEIKLKSSESVKLFYLDPAQKLDSSAEPKVSFEKFRKRIILYK 299
QY 371 CNDLSFNIQCCVAENEBEGPTNVBPFTYLSLFDIKYNNKISADPHVDLHPSVRQMIAT 430
300 CNDLSFNIQCCVAENEBEGPTNVBPFTYLSLFDIKYNNKISADPHVDLHPSVRQMIAT 359
QY 431 TSPALMNGSGPRTQALGIIHEAMQPKQISFVTCCHPDIPLVARIEKLGQSIITHC 490
360 TSPALMNGSG--OSPVLKGIHEAMQPKQISFVTCCHPDIPLVARIEKLGQSIITHC 418
QY 491 AEPYKSSDSSKVAQKVLKNAKQACORLQGYMPMAARLTFFKASGNDLNARFSAIY 550
419 AEPYKSSDSSKVAQKVLKNAKQACORLQGYMPMAARLTFFKASGNDLNARFSAIY 478
QY 551 RODSNKLSNDMLKLLADFRKPKKAKLPVILGNDITIDNVSSDPNNYVNSYIPTKQF 610
479 RODSNKLSNDMLKLLADFRKPKKAKLPVILGNDITIDNVSSDPNNYVNSYIPTKQF 538
QY 611 ETCSTPTTFEEVEEVPICPKHTOPTYTNHLVYYPKLYKXDSQKSPKANNIACIEF 670
539 ETCSTPTTFEEVEEVPICPKHTOPTYTNHLVYYPKLYKXDSQKSPKANNIACIEF 598
QY 671 KDSDEDSQPLKCIYGRPGGPVFTSAPAAVLHHHQNBEFYDEIKIELPTQLHEKHLL 730
599 KDSDEDSQPLKCIYGRPGGPVFTSAPAAVLHHHQNBEFYDEIKIELPTQLHEKHLL 658
QY 731 TFFHVSQDNSSKGSYTKKDDVETQVGYSWLPKDKGRVVTSEQHI PVSANLPSGYGYOE 790
659 TFFHVSQDNSSKGSYTKKDDVETQVGYSWLPKDKGRVVTSEQHI PVSANLPSGYGYOE 718
QY 791 LGMGRHYGPEIKWNGGKPLKISTHLVSTVYTOQHLLNFQYQCKTESGQAALGNELV 850
719 LGMGRHYGPEIKWNGGKPLKISTHLVSTVYTOQHLLNFQYQCKTESGQAALGNELV 778
QY 851 KYLKSILHMEGHVMIAPLPTIINQFLVLTTRATQEEVAVNVRVLIHHVAAQHEGLESJ 910
779 KYLKSILHMEGHVMIAPLPTIINQFLVLTTRATQEEVAVNVRVLIHHVAAQHEGLESJ 838
QY 911 LRSYKVAVKABPYVASEKTVHEELTKSMITILKPSADFLTSNKLRLRSWFEFFDLIKS 970
839 LRSYKVAVKABPYVASEKTVHEELTKSMITILKPSADFLTSNKLRLRSWFEFFDLIKS 898
QY 971 MAQHLIENSKVTLKNOQFPASVYHAAVETVNNMLMPHTQKQRNDPEASKNHNLSLAVPI 1030
899 MAQHLIENSKVTLKNOQFPASVYHAAVETVNNMLMPHTQKQRNDPEASKNHNLSLAVPI 958

```
Qy 1031 KRCFTMDRGVFKQJNNYISCFAPDPTLFEYKEFLRVYCNHEHYIPLNLPMPFGKG 1090
Db 959 KRCFTFMDRGVFKQJNNYISCFAPDPTLFEYKEFLRVYCNHEHYIPLNLPMPFGKG 1018
Qy 1091 RIQRVODLDVSLDDEFCRNHPLVGLLRVGTALQEFREYRLIAISVLKULLIKHSPD 1150
Db 1019 RIQRVODLDVSLDDEFCRNHPLVGLLRVGTALQEFREYRLIAISVLKULLIKHSPD 1078
Qy 1151 DRYASHOARIATLPLFGLLIENVORINRVDSPPVPMAGMTYKDESLPAVNPV 1210
Db 1079 DRYASHOARIATLPLFGLLIENVORINRVDSPPVPMAGMTYKDESLPAVNPV 1138
Qy 1211 TPQKSTLDSLHKDLGAIISGIASPYTSTPNINSVRNADSGSLISTDSGNSLPERNS 1270
Db 1139 TPQKSTLDSLHKDLGAIISGIASPYTSTPNINSVRNADSGSLISTDSGNSLPERNS 1198
Qy 1271 EKSNSLDKQOOSTLGNSVVRCDKLDQSEIKSLMCFYILKMSMDALFTYWNKASTSE 1330
Db 1199 EKSNSLDKQOOSTLGNSVVRCDKLDQSEIKSLMCFYILKMSMDALFTYWNKASTSE 1258
Qy 1331 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMMHARLQQLG 1367
Db 1255 LMDPFTISEVCLHOFQYMGKRYIARNQEGLPVHDKSCGTLPEVSNKRGMMHARLQQLG 1318
Qy 1368 SLDNSLTFNHSYGHSDADVLHOSLSEANATEVCLTALDTLSLFTLAFKQQLADHGNP 1427
Db 1319 SLDNSLTFNHSYGHSDADVLHOSLSEANATEVCLTALDTLSLFTLAFKQQLADHGNP 1378
Qy 1428 LMKKYFDVYLCFLQKHQOSETALKNVFTALRSILYKFPSTYEGRADMCALCYEILKCN 1487
Db 1379 LMKKYFDVYLCFLQKHQOSETALKNVFTALRSILYKFPSTYEGRADMCALCYEILKCN 1438
Qy 1488 SKLSSIRTEASOLVYLMNNNDYGGKSFVRTHLOVIVSGLADVIGIGTRPOQL 1547
Db 1439 SKLSSIRTEASOLVYLMNNNDYGGKSFVRTHLOVIVSGLADVIGIGTRPOQL 1498
Qy 1548 SIINNCAANDRLIKHTSFSSDVYDLTKRIRTVLMATAQKCEHNDPEMLVDIQLYSKSY 1607
Db 1499 SIINNCAANDRLIKHTSFSSDVYDLTKRIRTVLMATAQKCEHNDPEMLVDIQLYSKSY 1558
Qy 1608 ASTPELRKTWLDSMARIVYNGDLSBAMCYHVTALVAEYLTRKCAVOMEPPLPHSHS 1667
Db 1559 ASTPELRKTWLDSMARIVYNGDLSBAMCYHVTALVAEYLTRK----- 1603
Qy 1668 ACLRSRGVPRQGCAPRVTIPNIDEASMMEDVGMQVHFVEDVLMELLOCADGLMK 1727
Db 1604 -----GVPRQGCAPRVTIPNIDEASMMEDVGMQVHFVEDVLMELLOCADGLMK 1655
Qy 1728 AERYELADIYKLIIPYERKDFERLAHLYDLHRAYSKTEVMHSGRLLGTYFRVAF 1787
Db 1656 AERYELADIYKLIIPYERKDFERLAHLYDLHRAYSKTEVMHSGRLLGTYFRVAF 1678
Qy 1788 PQOAAQYOFPTDSETVGFFPEDEDEGKEYIYKEPKLTPLEISQRLIKLYSDKFGSENVK 1847
Db 1679 -----FFPEDEDEGKEYIYKEPKLTPLEISQRLIKLYSDKFGSENVK 1720
Qy 1848 IODSGKVNPKDLSKRAYIQVTHVLPFPEKELQEKTFPESHNRIRRMFEMPFTQTK 1907
Db 1721 IODSGKVNPKDLSKRAYIQVTHVLPFPEKELQEKTFPESHNRIRRMFEMPFTQTK 1780
Qy 1908 ROGGEQCKRRITILAHICFPYVKRIRPVMTQHTDMLPIEVAIDEMSKYVLAELQOLS 1967
Db 1781 ROGGEQCKRRITILAHICFPYVKRIRPVMTQHTDMLPIEVAIDEMSKYVLAELQOLS 1840
Qy 1968 SAEVMDIKLQLKQSSVQVNAAGPLAVARAFLDNTNRYRPNKYKLLKEYFRQFVAC 2027
Db 1841 SAEVMDIKLQLKQSSVQVNAAGPLAVARAFLDNTNRYRPNKYKLLKEYFRQFVAC 1900
Qy 2028 GQALVNEELIKEDOLEYOEEMKANYRMAKELSEIMHBOICPLEKTSVLPNSLIHFA 2087
Db 1901 GQALVNEELIKEDOLEYOEEMKANYRMAKELSEIMHBOICPLEKTSVLPNSLIHFA 1960
```

```
Qy 2088 ISGFTSTVHGMTSSSSSV 2107
Db 1961 ISGFTSTVHGMTSSSSSV 1980

RESULT 4
US-09-736-969A-94
; Sequence 94, Application US/09736969A
; Patent No. US20020068302A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-4 Transmembrane Protein
; FILE REFERENCE: 020054-000411US
; CURRENT APPLICATION NUMBER: US/09/736,969A
; PRIORITY FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 94
; LENGTH: 1980
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human CLASP-2
US-09-736-969A-94

Query Match 90.8%; Score 9926.5; DB 10; Length 1980;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1932; Conservative 8; Mismatches 15; Indels 105; Gaps 6;

Qy 73 MLFPYDPOGRIILRQGRYICSTYPAKAEBAOSLFTECKTYNSMHLVNYKEEDS 132
Db 1 MLFPYDPOGRIILRQGRYICSTYPAKAEBAOSLFTECKTYNSMHLVNYKEEDS 60

Qy 133 GEFROLPKRVVLDLCPVHYVYEVDEBVDKEDPAASLGSQKGIITGIGWLTKGNMNSAISV 192
Db 61 GEFROLPKRVVLDLCPVHYVYEVDEBVDKEDPAASLGSQKGIITGIGWLTKGNMNSAISV 120

Qy 193 TWRSFKRRFPFHILIQGDGSSYKFEFLKD--LQKEPKGSIFLGFYGVSPRNKVRFPAPFL 250
Db 121 TWRSFKRRFPFHILIQGDGSSYKFEFLKD--LQKEPKGSIFLGFYGVSPRNKVRFPAPFL 179
```

251 KNOOKSSYLLAADSEVEMERITILNKILQNEFAAMOEKRNDSHEDDEOSKLEGGSG 310
180 KNOOKSSYLLAADSEVEMERITILNKILQNEFAAMOEKRNDSHEDDEOSKLEGGSG 239
311 LQSYLELAKSAREABIKLKSESRYKLYLDPDAQKLDPSASPEVKSFEKKGKILVK 370
240 LQSYLELAKSAREABIKLKSESRYKLYLDPDAQKLDPSASPEVKSFEKKGKILVK 299
371 CNDLSNLOCCVAENEEGPTTNEPFFVTLSPFDIKYKRIKSNDSHVDLNHSVRQMT 430
300 CNDLSNLOCCVAENEEGPTTNEPFFVTLSPFDIKYKRIKSNDSHVDLNHSVRQMT 359
431 TSPALNNGSGPETQSALRGILHEAAMQYKOGIFSTVCPHDI.FLVARIEKVLQSGITHC 490
360 TSPALNNGSG-GSPVTLKGLHEAAMQYKOGIFSTVCPHDI.FLVARIEKVLQSGITHC 418
491 ABEYKSSDSKVAQKVLKNAKQACORLQYRMPFMAARTLEFKASGMLDKNARPSATY 550
419 ABEYKSSDSKVAQKVLKNAKQACORLQYRMPFMAARTLEFKASGMLDKNARPSATY 478
551 RQDSNKLSDMDMLKILADPRKPEKAKL.PVILGNLDTITDNVSDPPTVNSSYIPTKOF 610
479 RQDSNKLSDMDMLKILADPRKPEKAKL.PVILGNLDTITDNVSDPPTVNSSYIPTKOF 538
611 ETCSTKPTIFEEVEBFVPCIPKHQPYTIYTNHLVYVPKYLKQOSKSPAKARIALCIF 670
539 ETCSTKPTIFEEVEBFVPCIPKHQPYTIYTNHLVYVPKYLKQOSKSPAKARIALCIF 598
671 KQDESDQPLKCIYRPGGPFVTRSAFAVLAHHQNEPFDYKIELPTOLHEKHLIL 730
599 KQDESDQPLKCIYRPGGPFVTRSAFAVLAHHQNEPFDYKIELPTOLHEKHLIL 658
731 TEFHVSQDSSKSTKKRDVETQVGYSMPLPKDRVVTSEQHIEVSNLPGHLYG 790
659 TEFHVSQDSSKSTKKRDVETQVGYSMPLPKDRVVTSEQHIEVSNLPGHLYG 718
791 LGMGRHYGPEIKKVDGKPLKISTHVLSTVYQDOHNLNFQYCKTESGAGALNELY 850
719 LGMGRHYGPEIKKVDGKPLKISTHVLSTVYQDOHNLNFQYCKTESGAGALNELY 778
851 KYLKSIAHEGHWMI.FLPTILNQLFRVLTARATQEEVAVNVTRVILHVAQCHEBGLSH 910
779 KYLKSIAHEGHWMI.FLPTILNQLFRVLTARATQEEVAVNVTRVILHVAQCHEBGLSH 838
911 LBSYVYKAAKABRYVASEKTVHEELTKGMTILKRSADPLSNKLKXSWFFDVLIS 970
839 LBSYVYKAAKABRYVASEKTVHEELTKGMTILKRSADPLSNKLKXSWFFDVLIS 898
971 MAQHLIENSKVLLRNQRPASVYHAAVETVNM.LMHIQKFRDNPBASKNANHS.LAVFI 1030
899 MAQHLIENSKVLLRNQRPASVYHAAVETVNM.LMHIQKFRDNPBASKNANHS.LAVFI 958
1031 KQCFTEMDRGFVKQJNNYISCFAPGDPKTLFEYKFEPLRVVNCNHEHYI.PLNLPMDFGK 1090
959 KQCFTEMDRGFVKQJNNYISCFAPGDPKTLFEYKFEPLRVVNCNHEHYI.PLNLPMDFGK 1018
1091 RIORVODLQDLSLTDEFCRNHLVGLLREVTALQERREVLIAISVLKXLLHGSFD 1150
1019 RIORVODLQDLSLTDEFCRNHLVGLLREVTALQERREVLIAISVLKXLLHGSFD 1078
1151 DRYASRHOARIATLVLPLFGLLIENVORINVDVPPFNAGMTWKDES.LALPANNPLY 1210
1079 DRYASRHOARIATLVLPLFGLLIENVORINVDVPPFNAGMTWKDES.LALPANNPLY 1138
1211 TPQKSTLQNSLHKDLGAIISGIASPYTSTPINSVNRNDSRGLISIDSGNSLPERNS 1270
1139 TPQKSTLQNSLHKDLGAIISGIASPYTSTPINSVNRNDSRGLISIDSGNSLPERNS 1198
1271 EKSNSLQKXQSSSTLGNVVRCDKLDOSERKSLMLKELVILKMSMDALFTYWNKSTSE 1330
1199 EKSNSLQKXQSSSTLGNVVRCDKLDOSERKSLMLKELVILKMSMDALFTYWNKSTSE 1258

1331 LMDPFTISEVCLHOFQYMKRYIA-----RTGMHARLQOLG 1367
1259 LMDPFTISEVCLHOFQYMKRYIA-----RTGMHARLQOLG 1318
1368 SIDNSLTENHSGHSDADVLHOSLLEANIATEVCLTALDTS.LFTLAFKQOLLADHGNP 1427
1319 SIDNSLTENHSGHSDADVLHOSLLEANIATEVCLTALDTS.LFTLAFKQOLLADHGNP 1378
1428 LMKXVFDVYLCLOHQSETLKXVFTLRSLIYFSPSTFEGRADMCALCYELTKCN 1487
1379 LMKXVFDVYLCLOHQSETLKXVFTLRSLIYFSPSTFEGRADMCALCYELTKCN 1438
1488 SKLSSIRTEASQLFELRNNDYTGKSFVTHLQVLIISVQLIADVYIGSTFPOSL 1547
1439 SKLSSIRTEASQLFELRNNDYTGKSFVTHLQVLIISVQLIADVYIGSTFPOSL 1498
1548 SIINNCANDRLIKHTSFSSDVKDLTKRIRVLYMATQMKHENDPEMLVDLOYSLASY 1607
1439 SIINNCANDRLIKHTSFSSDVKDLTKRIRVLYMATQMKHENDPEMLVDLOYSLASY 1558
1608 ASTPELRKTLQDSMARIVHKNQDLSSEAMCYVHTALVAEYLTREAVQWEPRLPHSHS 1667
1559 ASTPELRKTLQDSMARIVHKNQDLSSEAMCYVHTALVAEYLTREAVQWEPRLPHSHS 1603
1668 ACLRRSGGVFROGCTAFRVITPNIDEASNMEDVGMODVHFNEDVLMELLEQCADGLMK 1727
1604 -----GVFRQGTAFRVITPNIDEASNMEDVGMODVHFNEDVLMELLEQCADGLMK 1655
1728 AERYELADIYKLIPIYERKRDPERLAHLVDTLHRAVSKTEVHSGRRLIGTYFRVAF 1787
1656 AERYELADIYKLIPIYERKRD----- 1678
1788 FQGAHQYQFQSETEVEGFEDEDEKXYIYKPKLTP.LSEISORLLKLYSDKFGSENVKM 1847
1679 -----FPEDEDEKXYIYKPKLTP.LSEISORLLKLYSDKFGSENVKM 1720
1848 IODSGKVPKQDLSKXAYIOVTHVLPFDEKELQERKTEPERSHNIIRFMEPMTOTGK 1907
1721 IODSGKVPKQDLSKXAYIOVTHVLPFDEKELQERKTEPERSHNIIRFMEPMTOTGK 1780
1908 ROGVEBOCKRTIITAIHCPFYVKRI.PWYQHHTDNLPIEVAIDEMSKYAEI.ROLCS 1967
1781 ROGVEBOCKRTIITAIHCPFYVKRI.PWYQHHTDNLPIEVAIDEMSKYAEI.ROLCS 1840
1968 SAEVDMIKLOQLQGSVSQVNAAGLAPARAF.LDNTKRYVDNRYKLLKEVFRQFVEAC 2027
1841 SAEVDMIKLOQLQGSVSQVNAAGLAPARAF.LDNTKRYVDNRYKLLKEVFRQFVEAC 1900
2028 GQALAVNERLIKEDOLEYQOEKKNARYREMAKELSEIMHEQICPLEKTSVL.PNSLHIFNA 2087
1901 GQALAVNERLIKEDOLEYQOEKKNARYREMAKELSEIMHEQICPLEKTSVL.PNSLHIFNA 1960
2088 ISGTPSTVMHGMSSSVV 2107
1961 ISGTPSTVMHGMSSSVV 1980

RESULT 5
US-09-736-960-91
; Sequence 91, Application US/09736960
; Patent No. US20020102267A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; TITLE OF INVENTION: CLASP-5 Transmembrane Protein
; FILE REFERENCE: 020054-00051IUS
; CURRENT APPLICATION NUMBER: US/09/736,960
; PRIOR FILING DATE: 2001-09-20
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498

/ PRIOR FILING DATE: 1999-10-29
/ PRIOR APPLICATION NUMBER: US 60/170,453
/ PRIOR FILING DATE: 1999-12-13
/ PRIOR APPLICATION NUMBER: US 60/176,195
/ PRIOR FILING DATE: 2000-01-14
/ PRIOR APPLICATION NUMBER: US 60/182,296
/ PRIOR FILING DATE: 2000-02-14
/ PRIOR APPLICATION NUMBER: US 09/547,276
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,267
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,460
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,527
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,528
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 09/687,837
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,503
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,508
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,539
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,543
/ PRIOR FILING DATE: 2000-10-13
/ NUMBER OF SEQ ID NOS: 134
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 91
/ LENGTH: 1980
/ TYPE: PRF
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: human CLASP-2
US-09-736-960-91

Query Match 90.8%; Score 9926.5; DB 10; Length 1980;
Best Local Similarity 93.8%; Pred. No. 0;
Matches 1992; Conservative 8; Mismatches 15; Indels 105; Gaps 6;

73 MLFPYDFQTAALRQGRYICSTVPKABEBAQSLFVTECTKTNSDMLVNYKEDYS 132
1 MLFPYDFQTAALRQGRYICSTVPKABEBAQSLFVTECTKTNSDMLVNYKEDYS 60
133 GEFROLPNKRVKLDKLPVHYVEVDEVEDDEDAASIGSGKGTITKGMILYKGNMSAISV 192
61 GEFROLPNKRVKLDKLPVHYVEVDEVEDDEDAASIGSGKGTITKGMILYKGNMSAISV 120
193 TRRSFRRFFHLIQLDGSYKPEFLKD--LOKEPKSIFLGFLYGVSPNNKVRPAPFEL 250
121 TRRSFRRFFHLIQLDGSYKPEFLKD--LOKEPKSIFLGFLYGVSPNNKVRPAPFEL 179
251 KMOXSSYLLAADSEVEMEWITILNKILQLNFEAAMQIKRNGDSHEDDEQSLDESGSG 310
180 KMOXSSYLLAADSEVEMEWITILNKILQLNFEAAMQIKRNGDSHEDDEQSLDESGSG 239
311 LQSYLPELAKSAREAEIKLKSSRYVLFTLPDPAQCLDSSAEPYKSSFEKRGKRIIVK 370
240 LQSYLPELAKSAREAEIKLKSSRYVLFTLPDPAQCLDSSAEPYKSSFEKRGKRIIVK 299
371 CNDLSFNLOCCVAENEGPFTTVNEPFTVLSLFDIKYNNKISADPFVNDLHFSVRQMIAT 430
300 CNDLSFNLOCCVAENEGPFTTVNEPFTVLSLFDIKYNNKISADPFVNDLHFSVRQMIAT 359
431 TSPALMNGSGPETOSALRGILHEAAMQPKQIFSVTCCHPDIFLVARIEKVLQGSITHC 490
360 TSPALMNGSG--QSPSYLKGILHEAAMQPKQIFSVTCCHPDIFLVARIEKVLQGSITHC 418
491 ASPYKSSSRYAQCQVULKNAKQAQORLGOYRMPFAMARLTPKDSASGULDKNARPSATY 550
419 ASPYKSSSRYAQCQVULKNAKQAQORLGOYRMPFAMARLTPKDSASGULDKNARPSATY 478

QY 551 RODSNKLSNDMDKLADFRKPERKAKLPEVILGNLDTITDYNVSDFPNVNSSYIPTKQF 610
DB 479 RODSNKLSNDMDKLADFRKPERKAKLPEVILGNLDTITDYNVSDFPNVNSSYIPTKQF 538
QY 611 ETCSKTPTTREVEEVPCIPKATQPTTYTNHLVYYPKYLKDSQKSPKANINACIEF 670
DB 539 ETCSKTPTTREVEEVPCIPKATQPTTYTNHLVYYPKYLKDSQKSPKANINACIEF 598
QY 671 KDSDEDSOPLKCIIGRPGCPVFTSAPFAVAHNNQPEYDIEKLEPTOLHEKHHLL 730
DB 599 KDSDEDSOPLKCIIGRPGCPVFTSAPFAVAHNNQPEYDIEKLEPTOLHEKHHLL 658
QY 731 TFFHVSQNSKSGSTRKDDVETQVGSWLPPLKQGRVVSQOHIPVSNLPSGLGYOE 790
DB 659 TFFHVSQNSKSGSTRKDDVETQVGSWLPPLKQGRVVSQOHIPVSNLPSGLGYOE 718
QY 791 LQNGHYGPEIKVNDGKPLKISTHLYSVTVDQOHLNFPQYCKTESGAQALGNEV 850
DB 739 LQNGHYGPEIKVNDGKPLKISTHLYSVTVDQOHLNFPQYCKTESGAQALGNEV 778
QY 851 KYLKSILHMEGHVMTAFLPTILNOLFRVLTRATQSEVAVNVRVLIHVAQCHBGLBSH 910
DB 779 KYLKSILHMEGHVMTAFLPTILNOLFRVLTRATQSEVAVNVRVLIHVAQCHBGLBSH 838
QY 911 LRSYKAYKAPYVYASEKYVHEELTKSMITILKPSADFLTSNKLKYSWFFPDVLIS 970
DB 839 LRSYKAYKAPYVYASEKYVHEELTKSMITILKPSADFLTSNKLKYSWFFPDVLIS 898
QY 971 MAQHLIENSKVLLNORFPASYNHAAVETVNMMLPHITQKRDNPBASKANHSIAV 1030
DB 899 MAQHLIENSKVLLNORFPASYNHAAVETVNMMLPHITQKRDNPBASKANHSIAV 958
QY 1031 KRCPFMGRGPFVKOINNYISCFAPGDPKTLLEYKPEFLRVVNCNHHYPLNLPMPFGG 1090
DB 959 KRCPFMGRGPFVKOINNYISCFAPGDPKTLLEYKPEFLRVVNCNHHYPLNLPMPFGG 1018
QY 1091 RIQRYODLQDYSLTDEFCRNHFLVGLLREVGTALQEBREVRLLAISVLKNLLIKHSPD 1150
DB 1019 RIQRYODLQDYSLTDEFCRNHFLVGLLREVGTALQEBREVRLLAISVLKNLLIKHSPD 1078
QY 1151 DRYASRSHQARIATYLPFLGILLIENVQRIINRDVSPFPVNGMTYKDESLALPAVNPV 1210
DB 1079 DRYASRSHQARIATYLPFLGILLIENVQRIINRDVSPFPVNGMTYKDESLALPAVNPV 1138
QY 1211 TPQKSTLQNSLHKQLLGAISGIAAPYTTSTPNINSVRNADRSGLISTDSNSLPERNS 1270
DB 1139 TPQKSTLQNSLHKQLLGAISGIAAPYTTSTPNINSVRNADRSGLISTDSNSLPERNS 1198
QY 1271 EKSNSLDKHQOSSLTGNNSVVRCDKLDQSEIKSLMCFYIILKSMGSDDALFTYNNKASTSE 1330
DB 1199 EKSNSLDKHQOSSLTGNNSVVRCDKLDQSEIKSLMCFYIILKSMGSDDALFTYNNKASTSE 1258
QY 1331 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1367
DB 1259 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1318
QY 1367 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1427
DB 1339 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1378
QY 1427 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1487
DB 1379 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1438
QY 1487 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1547
DB 1439 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1498
QY 1547 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1607
DB 1499 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1558
QY 1607 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1667
DB 1558 LMDPFTISEVCLHOFQYMGKRYIA-----RTGMHARLQOLG 1617

```
|||||
Db 1559 ASTPELRKTMIDSMARHVNKGDLSBAACYYHVTALVAEYLTRK----- 1603
1668 ACIARRRGVRCGCGTAFRITPNIDEBSMMDVGMODVHEMEDVLMELLECCADLWK 1727
Db 1604 -----GVFRGGCTAFRITTNIDEBSMMDVGMODVHEMEDVLMELLECCADLWK 1655
1728 AERVELIADYKLIIFIEKRDPERLALHYDTLHRAYSKVTEVMHSGRRLIGTYFRVAF 1787
Db 1656 AERVELIADYKLIIFIEKRD----- 1678
1788 FGGAAOYQFTDSETDVEGFEDEDGKEYIYKEPKLTPLSEISQRLIKLYSDKFGSNVAK 1847
Db 1679 -----FPEDEDGKEYIYKEPKLTPLSEISQRLIKLYSDKFGSNVAK 1720
1848 IODSGKVPKDLDSKAYIQVTHVIPFPEKELOEKTEPERSHNIRRFEMEPFOTOK 1907
Db 1721 IODSGKVPKDLDSKAYIQVTHVIPFPEKELOEKTEPERSHNIRRFEMEPFOTOK 1780
1908 RCGVEEQCKRRTILTAHCFPVYKRIPIWYQHTDLPPIEVAIDEMSKVAELRQLCS 1967
Db 1781 RCGVEEQCKRRTILTAHCFPVYKRIPIWYQHTDLPPIEVAIDEMSKVAELRQLCS 1840
1968 SAEVDMIKQLKIQSSVSQVNAAPLAAYAFLLDNTNRYPDNKVKLLKEVFRQVEAC 2027
Db 1841 SAEVDMIKQLKIQSSVSQVNAAPLAAYAFLLDNTNRYPDNKVKLLKEVFRQVEAC 1900
2028 GALLANERLIKEDOLEYOEMKANYREMAKELSEIMHGOICPLEKTSVLPSLHIFNA 2087
Db 1901 GALLANERLIKEDOLEYOEMKANYREMAKELSEIMHGOICPLEKTSVLPSLHIFNA 1960
2088 ISGTPTSTWVHGMTSSSSSVY 2107
Db 1961 ISGTPTSTWVHGMTSSSSSVY 1980

RESULT 6
US-09-978-244A-27
; Sequence 27, Application US/09978244A
; Publication No. US20030103992A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter S
; APPLICANT: Garman, Jonathan D.
; APPLICANT: Candia III, Albert F.
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP MEMBRANE PROTEINS
; FILE REFERENCE: 020554-000161US
; CURRENT APPLICATION NUMBER: US/09/978, 244A
; PRIOR FILING DATE: 2001-10-15
; PRIOR APPLICATION NUMBER: US 60/310,028
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 09/737,246
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/736,969
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/736,960
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 09/736,968
; PRIOR FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/240,545
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; Remaining Prior Application data removed - See file wrapper or PALM.
; NUMBER OF SEQ ID NOS: 106
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
```

```
; LENGTH: 2040
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: CLASP-2
US-09-978-244A-27

Query Match 90.4%; Score 9886.5; DB 9; Length 2040;
Best Local Similarity 91.1%; Pred. No. 0;
Matches 1932; Conservative 8; Mismatches 15; Indels 165; Gaps 7;

1 MLLPFYDDFQIALRRQGRYICSTVPAAAEBAQSLFVTECIKTYSNDHLVYKYEYDS 60
73 MLLPFYDDFQIALRRQGRYICSTVPAAAEBAQSLFVTECIKTYSNDHLVYKYEYDS 110
111 -----TECIKTYSNDHLVYKYEYDS 132
61 GEFROLPNKVPKLDKLPVHVEVDEADKEDDAASLGSTECIKTYNSDHLVYKYEYDS 120
133 GEFROLPNKVPKLDKLPVHVEVDEADKEDDAASLGSTECIKTYNSDHLVYKYEYDS 192
121 GEFROLPNKVPKLDKLPVHVEVDEADKEDDAASLGSTECIKTYNSDHLVYKYEYDS 180
193 TMRSEKRRPFHLIQLGDSYKFEPLKD--LQKPKGSIFLGFLYGVSPRNNKVRPAPEL 250
181 TMRSEKRRPFHLIQLGDSYKFEPLKD--LQKPKGSIFLGFLYGVSPRNNKVRPAPEL 239
251 KMODSSYLLADSEVMEEMWITLNLKILQNFEAAMOKRNGDSDHEDOSKLGSGSG 310
240 KMODSSYLLADSEVMEEMWITLNLKILQNFEAAMOKRNGDSDHEDOSKLGSGSG 299
311 LDSYLPBLAKSAREAEIKLKSSESVKLFYLDPAQKLPSSAPPEVKSFEKFGKRIYK 370
300 LDSYLPBLAKSAREAEIKLKSSESVKLFYLDPAQKLPSSAPPEVKSFEKFGKRIYK 359
371 CNDLSFNIOCCVAENBEGFTNVBPFTVLSLFDIKYRKISADPHVDLHNSVRQMIT 430
360 CNDLSFNIOCCVAENBEGFTNVBPFTVLSLFDIKYRKISADPHVDLHNSVRQMIT 419
431 TSPALMNGSGPETSALGILHEAMQYPKQISFVTCCHPPIFLVARIEKLGQSTHIC 490
420 TSPALMNGSG-OSPEVLKILHEAMQYPKQISFVTCCHPPIFLVARIEKLGQSTHIC 478
491 ABPYMKSSDSKVAOKVLAQAKACORLGOYMPFAMAARTLFDKASGLDNKARFSAIY 550
479 ABPYMKSSDSKVAOKVLAQAKACORLGOYMPFAMAARTLFDKASGLDNKARFSAIY 538
551 RQDSNKLSDMDLKLADFRKPEKRAKLPVIGNDITIDNVSDFPNVNSSYIPTKQF 610
539 RQDSNKLSDMDLKLADFRKPEKRAKLPVIGNDITIDNVSDFPNVNSSYIPTKQF 598
611 ETCSKTPITFEVEEVPICPKHTOPYTYTNHLYYPRYLKVDQSKSPKAKANINACIEF 670
599 ETCSKTPITFEVEEVPICPKHTOPYTYTNHLYYPRYLKVDQSKSPKAKANINACIEF 658
671 KQDSEDSQPLKCIYGRPGPVFTSAPAAVLAHHNQNEFYDEIKIELPTQHEKHLHL 730
659 KQDSEDSQPLKCIYGRPGPVFTSAPAAVLAHHNQNEFYDEIKIELPTQHEKHLHL 718
731 TFFHVSNDNSKSGSTKRDVETQVGYSLPLKDGKRVVTSQHLIVSANLPSGYLQOE 790
719 TFFHVSNDNSKSGSTKRDVETQVGYSLPLKDGKRVVTSQHLIVSANLPSGYLQOE 778
791 LQMGHRYGPEIKWVGGRPLKISTHLVSTVYTOQHLLNFQYQKTESGQALGNELV 838
779 LQMGHRYGPEIKWVGGRPLKISTHLVSTVYTOQHLLNFQYQKTESGQALGNELV 838
851 KTLKSLHAMEGHVMIAPPTILNQFLVLTTRATQEVAVNVTYRVIHHVAQHEGEGLESH 910
839 KTLKSLHAMEGHVMIAPPTILNQFLVLTTRATQEVAVNVTYRVIHHVAQHEGEGLESH 898
911 LRSYKAVAKAPYVASEYKTVHEELTKSMITLKPADPLTSNKLKYSNFFDVLIKS 970
```

```

Db 899 LRSYKAYKAEPPYVASEYKTVHEBLTKSMITTLKPSADPLTSNKLRLRSWFFDVLKLS 958
Qy 971 MAOHLIENSKVTLRLNORPPASVHVAETVVMMLMPLHTQKFRDNEPASKNHLAVPI 1030
Db 959 MAOHLIENSKVTLRLNORPPASVHVAETVVMMLMPLHTQKFRDNEPASKNHLAVPI 1018
Qy 1031 KRCFTFMDRGPFVKQJNNYISCPAPDPTLFEYKEFLRVYCNHEHYIPLMLPMPFGKG 1090
Db 1019 KRCFTFMDRGPFVKQJNNYISCPAPDPTLFEYKEFLRVYCNHEHYIPLMLPMPFGKG 1078
Qy 1091 RIQRVODLDVSLDEFCRNHPVLGGLREVGTALQEPREVRLLAISVTLKULLIKHSD 1150
Db 1079 RIQRVODLDVSLDEFCRNHPVLGGLREVGTALQEPREVRLLAISVTLKULLIKHSD 1138
Qy 1151 DRYASRSHOARIATLTLPLFGLLIEVORINRDVSPFPVNAGMTVKDESLAPVNPV 1210
Db 1139 DRYASRSHOARIATLTLPLFGLLIEVORINRDVSPFPVNAGMTVKDESLAPVNPV 1198
Qy 1211 TPQKSTLDNSLHKDLGALSGIASPYTTSTPNINSVRNADSRGSLISTSGNSLPERNS 1270
Db 1199 TPQKSTLDNSLHKDLGALSGIASPYTTSTPNINSVRNADSRGSLISTSGNSLPERNS 1258
Qy 1271 EKSNSLDRKQOSTLGNSVVRCDKLDQSEIKSLMCFYTLKMSDDALFTYWKASTSE 1330
Db 1259 EKSNSLDRKQOSTLGNSVVRCDKLDQSEIKSLMCFYTLKMSDDALFTYWKASTSE 1318
Qy 1331 LMDPFTISFVCLHOPQYMGKRYIA-----RTGMHARLQOLG 1367
Db 1319 LMDPFTISFVCLHOPQYMGKRYIARNOBGLGPIVHDKSGQTLPVSRNRTGMHARLQOLG 1378
Qy 1368 SLDNSLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDLSLFTLAFNQLADHGNP 1427
Db 1379 SLDNSLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDLSLFTLAFNQLADHGNP 1438
Qy 1428 LMKKVPDVLCTLOKQOSTALKNVFTALRSLLYKPESTFYGRADMCALCEYILKCN 1487
Db 1439 LMKKVPDVLCTLOKQOSTALKNVFTALRSLLYKPESTFYGRADMCALCEYILKCN 1498
Qy 1488 SKLSSIRTEASOLVFLMKNNDPYGKSFVRTHLOVILSVSOLADVVIGIGSTRQOGL 1547
Db 1499 SKLSSIRTEASOLVFLMKNNDPYGKSFVRTHLOVILSVSOLADVVIGIGSTRQOGL 1558
Qy 1548 SIINNCAANDRLIKHTSFSSDVLTKRIRTVLMATAQKHEHNDPEMLVDLOYSLAKSY 1607
Db 1559 SIINNCAANDRLIKHTSFSSDVLTKRIRTVLMATAQKHEHNDPEMLVDLOYSLAKSY 1618
Qy 1608 ASTPELRKTWLDSMARIHYNGDLSBANACYHTALVAEYLTRKEAVMBEPLLPKSHS 1667
Db 1639 ASTPELRKTWLDSMARIHYNGDLSBANACYHTALVAEYLTRK----- 1663
Qy 1668 ACLARSRGVPRQCGTAPRVITPNIDEASMMEDVGMQVHNFEDVLMELBQCADGLMK 1727
Db 1664 -----GVPRQCGTAPRVITPNIDEASMMEDVGMQVHNFEDVLMELBQCADGLMK 1715
Qy 1728 AERYELIADLYKLIIPYKERRDPERLAHLYDLTHRAVEKTVEMHSGRLLGYFRVAF 1787
Db 1716 AERYELIADLYKLIIPYKERRD----- 1738
Qy 1788 PQQAAOYQFTDSETVDEGFPEDDEGKEYIYKEPKLTPLEISORLIKLYSDKFGSENVMK 1847
Db 1739 -----PFEDEDEGKEYIYKEPKLTPLEISORLIKLYSDKFGSENVMK 1780
Qy 1848 IODSGKVNPKDLDSKAYAIQVHVLPFPPEKELQEKTKFSBNSHNRRMFEMPFTQTK 1907
Db 1781 IODSGKVNPKDLDSKAYAIQVHVLPFPPEKELQEKTKFSBNSHNRRMFEMPFTQTK 1840
Qy 1908 RQGGVEEQCKRRTILTAHCPCPVYKRIIPVMTQHTDNLPIEVALIDEMSKVAELRQLCS 1967
Db 1841 RQGGVEEQCKRRTILTAHCPCPVYKRIIPVMTQHTDNLPIEVALIDEMSKVAELRQLCS 1900
Qy 1968 SAEVDMIKQLCKQSVSVQVNAAGPLAYARAFLDNTNTRYKYPDNKYLKEYFRQVENC 2027
Db 1901 SAEVDMIKQLCKQSVSVQVNAAGPLAYARAFLDNTNTRYKYPDNKYLKEYFRQVENC 1960

```

```

Qy 2028 GOALAVNERLLIKEDQLEYQEBMKANYREMAKELSEIMHEQICPLEEKTSVLPNSLHIFNA 2087
Db 1961 GOALAVNERLLIKEDQLEYQEBMKANYREMAKELSEIMHEQICPLEEKTSVLPNSLHIFNA 2020
Qy 2088 ISGTPTSTVHGHTSSSSSV 2107
Db 2021 ISGTPTSTVHGHTSSSSSV 2040

RESULT 7
US-09-736-968A-10
/ Sequence 10, Application US/09736968A
/ Patent No. US20020169283A1
/ GENERAL INFORMATION:
/ APPLICANT: Lu, Peter
/ APPLICANT: Garman, Jonathan David
/ APPLICANT: Candia III, Albert Frederick
/ APPLICANT: Arbor Vita Corporation
/ TITLE OF INVENTION: CLASP-7 Transmembrane Protein
/ FILE REFERENCE: 020054-000611US
/ CURRENT APPLICATION NUMBER: US/09/736,968A
/ CURRENT FILING DATE: 2000-12-13
/ PRIOR APPLICATION NUMBER: US 60/160,860
/ PRIOR FILING DATE: 1999-10-21
/ PRIOR APPLICATION NUMBER: US 60/162,498
/ PRIOR FILING DATE: 1999-10-29
/ PRIOR APPLICATION NUMBER: US 60/170,453
/ PRIOR FILING DATE: 1999-12-13
/ PRIOR APPLICATION NUMBER: US 60/176,195
/ PRIOR FILING DATE: 2000-01-14
/ PRIOR APPLICATION NUMBER: US 60/182,296
/ PRIOR FILING DATE: 2000-02-14
/ PRIOR APPLICATION NUMBER: US 09/547,276
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,267
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,460
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,527
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 60/196,528
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 09/687,837
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,503
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,508
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,539
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,543
/ PRIOR FILING DATE: 2000-10-13
/ SOFTWARE: PatentIn Ver. 2.1
/ NUMBER OF SEQ ID NOS: 115
/ SEQ ID NO 10
/ LENGTH: 1534
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: human KIAA1058
US-09-736-968A-10

Query Match 72.9%; Score 7976; DB 9; Length 1534;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 536 ASGNLDKARFSATYRQDSNKLSDNDMLKLADEFKPKMAKLPLYLGNLDTIDNVSDD 595
Db 1 ASGNLDKARFSATYRQDSNKLSDNDMLKLADEFKPKMAKLPLYLGNLDTIDNVSDD 60
Qy 596 PPNVNSSYIPTKQETCKPTTFEEVEFPVCPJPKHTQPYTIYTNHLVYVPKYKYSQ 655

```

Db 61 PBYVNVSSYIPTKOFETCSKTPITFEVEBEVPCIPRHOTOPYITYNHLYVYPKLYKQSDQ 120
Qy 656 KSPAKARNIAICIEFDOSDEBDSQPLKCIYGRPGGVFTRSAPAAVLHHHONBEYDEIK 715
Db 121 KSPAKARNIAICIEFDOSDEBDSQPLKCIYGRPGGVFTRSAPAAVLHHHONBEYDEIK 180
Qy 716 IELPTQLEHKNHLLTFPHVSCDSSKSGSTKKRDVETOVGVSWPLKXGRVVTSEQHI 775
Db 181 IELPTQLEHKNHLLTFPHVSCDSSKSGSTKKRDVETOVGVSWPLKXGRVVTSEQHI 240
Qy 776 PVSANIPSGYLQELGNGRHYPETIKMDGKPLKISTHLVSTYTYTODOLHNFQYC 835
Db 241 PVSANIPSGYLQELGNGRHYPETIKMDGKPLKISTHLVSTYTYTODOLHNFQYC 300
Qy 836 OKTESGAOLGNELVYKLSLHMEGHVMIAPLTLINOLFRVLTATOEVAVANTRYI 895
Db 301 OKTESGAOLGNELVYKLSLHMEGHVMIAPLTLINOLFRVLTATOEVAVANTRYI 360
Qy 896 IHVVAOCHEEGLESNLSRYKVAAYKAPYVASEYKTVEBELTKSMITLIKPSADFLTSNK 955
Db 361 IHVVAOCHEEGLESNLSRYKVAAYKAPYVASEYKTVEBELTKSMITLIKPSADFLTSNK 420
Qy 956 LKYSWFFPDVULIKSAOHLIENSKYKLRNORFPASSTYHAAVETVYNNMLPHITQKFRDN 1015
Db 421 LKYSWFFPDVULIKSAOHLIENSKYKLRNORFPASSTYHAAVETVYNNMLPHITQKFRDN 480
Qy 1016 PEASKRANSLAVFIKRCFTFMDRGVFPKQINNYISCFAPGDPDKTLFEYKFEFLRVVGNH 1075
Db 481 PEASKRANSLAVFIKRCFTFMDRGVFPKQINNYISCFAPGDPDKTLFEYKFEFLRVVGNH 540
Qy 1076 EHYIPLNLEMPFGKRIQRYODLQDLSYLTDEFGRNHLVGLLREBVGALOEFRREVLI 1135
Db 541 EHYIPLNLEMPFGKRIQRYODLQDLSYLTDEFGRNHLVGLLREBVGALOEFRREVLI 600
Qy 1136 A1SVLKNLLIKHSFDRYASRSHQARIATLYPLFELLLENVQRIWVRVSPPVNAGMT 1195
Db 601 A1SVLKNLLIKHSFDRYASRSHQARIATLYPLFELLLENVQRIWVRVSPPVNAGMT 660
Qy 1196 VDESEALPAVNPVLPKSGSTLDNSLHKLGAISGASPYTSTSPNINSVRNADSRGS 1255
Db 661 VDESEALPAVNPVLPKSGSTLDNSLHKLGAISGASPYTSTSPNINSVRNADSRGS 720
Qy 1256 LISTDSGNSLPERNSEKNSLDRHQSSSTLGNSVVRCDKLDOSEIKSLMCELYILKSN 1315
Db 721 LISTDSGNSLPERNSEKNSLDRHQSSSTLGNSVVRCDKLDOSEIKSLMCELYILKSN 780
Qy 1316 DDALFTYNNKASISELMDFFTISEVCLHOFQYWGKRYIARTGMARLOOLGSLDNLSTF 1375
Db 781 DDALFTYNNKASISELMDFFTISEVCLHOFQYWGKRYIARTGMARLOOLGSLDNLSTF 840
Qy 1376 NNSYGHSDADVHLQSLLEANIATEVCLTALDTLSLFTLAFKQQLADHGNPLMKKVPDY 1435
Db 841 NNSYGHSDADVHLQSLLEANIATEVCLTALDTLSLFTLAFKQQLADHGNPLMKKVPDY 900
Qy 1436 YLCFLOKHQSEFALKNVFTALNSLIYKPSSTYEGRADMCALCYEILKCNCKLSISRT 1495
Db 901 YLCFLOKHQSEFALKNVFTALNSLIYKPSSTYEGRADMCALCYEILKCNCKLSISRT 960
Qy 1496 EASOLLYFLMRNPNFDYTGKKSFRTHLOYIISVSOLIAVVGIGTRFOOSLSIINNCA 1555
Db 961 EASOLLYFLMRNPNFDYTGKKSFRTHLOYIISVSOLIAVVGIGTRFOOSLSIINNCA 1020
Qy 1556 SDRLIKHTSFSSDVKDLTKRIRTVLMAATOMEKHEHDPMLVDLOYSLAKSYASPELAK 1615
Db 1021 SDRLIKHTSFSSDVKDLTKRIRTVLMAATOMEKHEHDPMLVDLOYSLAKSYASPELAK 1080
Qy 1616 TWLDSMARIVXNGDLSSEAMCYVHTALVAEYLTRKEAVONPEPLPHSHSACLRSRG 1675
Db 1081 TWLDSMARIVXNGDLSSEAMCYVHTALVAEYLTRKEAVONPEPLPHSHSACLRSRG 1140
Qy 1676 GVPFROCTAFRVTTPNIDEEASMEDVGNQDVHFNEDVIMELLBOCADGLMAERYELIA 1735
Db 1141 GVPFROCTAFRVTTPNIDEEASMEDVGNQDVHFNEDVIMELLBOCADGLMAERYELIA 1200

Qy 1736 DIYKLIPIYEKRRDPERLALYDTLHRAYSKVEWMSGRRLLCTYFRVAFPGQAQOY 1795
Db 1201 DIYKLIPIYEKRRDPERLALYDTLHRAYSKVEWMSGRRLLCTYFRVAFPGQAQOY 1260
Qy 1796 FTDSSTVEGPFEDDEGKEYIYKEKPLTPELSEISQRLKLKYSDKGSSENVKMIODSGKVN 1855
Db 1261 FTDSSTVEGPFEDDEGKEYIYKEKPLTPELSEISQRLKLKYSDKGSSENVKMIODSGKVN 1320
Qy 1856 PKDLDSKAYIOWTHVIPPPEKELQERKTEPERSHNRFPWFEMPFTOTGKROGVBEQ 1915
Db 1321 PKDLDSKAYIOWTHVIPPPEKELQERKTEPERSHNRFPWFEMPFTOTGKROGVBEQ 1380
Qy 1916 CKRRTILTAHCFPVYKRIIPMYOHTDNLPIEVAIDEMSKVAELROLCSADEVMIK 1975
Db 1381 CKRRTILTAHCFPVYKRIIPMYOHTDNLPIEVAIDEMSKVAELROLCSADEVMIK 1440
Qy 1976 LQKLQGSVSVQVNAAGPLAYARAFDDTNTKRYPDNKKVLLKEVRFQVEACGQALAVNE 2035
Db 1441 LQKLQGSVSVQVNAAGPLAYARAFDDTNTKRYPDNKKVLLKEVRFQVEACGQALAVNE 1500
Qy 2036 RLKEDOLEYOEMKANYREMAKELSEIMHOI 2068
Db 1501 RLKEDOLEYOEMKANYREMAKELSEIMHOI 1533

RESULT 8
US-09-736-969A-10
Sequence 10, Application US/09736969A
Patent No. US2002068302A1
GENERAL INFORMATION:
APPLICANT: Lu, Peter
APPLICANT: Garman, Jonathan David
APPLICANT: Candia III, Albert Frederick
APPLICANT: Arbor Vita Corporation
TITLE OF INVENTION: CLASP-4 Transmembrane Protein
FILE REFERENCE: 020054-000411US
CURRENT APPLICATION NUMBER: US/09/736,969A
CURRENT FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 60/160,860
PRIOR FILING DATE: 1999-10-21
PRIOR APPLICATION NUMBER: US 60/162,498
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: US 60/170,453
PRIOR FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: US 60/176,195
PRIOR FILING DATE: 2000-01-14
PRIOR APPLICATION NUMBER: US 60/182,296
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: US 09/547,276
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,267
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,460
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,527
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 09/687,837
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,503
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,508
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,539
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,543
PRIOR FILING DATE: 2000-10-13
NUMBER OF SEQ ID NOS: 153
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 1534

```

/ TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: KIAA1058
;
US-09-736-969A-10

```

Query Match	72.9%	Score 7976;	DB 10;	Length 1534;
Best Local Similarity	99.9%	Pred. No. 0;		
Matches 1532; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0

Qy	536	ASGNLDKARFSAI	I	ROOSNKL	SNDML	CLADFKR	BEKMA	PLV	GLMD	DTID	VSSD	595
Db	1	ASGNLDKARFSAI	I	RODSNKL	SNDML	CLADFKR	BEKMA	PLV	GLMD	DTID	VSSD	60
Qy	596	PPNYNSSYI	I	PTKOPETS	KTP	I	TEVEBE	FV	CP	IKPH	TO	655
Db	61	PPNYNSSYI	I	PTKOPETS	KTP	I	TEVEBE	FV	CP	IKPH	TO	120
Qy	656	KSPAKAR	IAI	CIEFKD	DEDSO	PLKCI	YSG	PGG	V	FRSA	PAV	715
Db	121	KSPAAR	IAI	CIEFKD	DEDSO	PLKCI	YSG	PGG	V	FRSA	PAV	180
Qy	716	IELPTOL	HEKH	HL	LT	FFHV	CDNS	KGST	KRD	V	ETO	775
Db	181	IELPTOL	HEKH	HL	LT	FFHV	CDNS	KGST	KRD	V	ETO	240
Qy	776	PVSANL	PSGY	LG	YQEL	GMGR	HY	PEI	KW	DG	GK	835
Db	241	PVSANL	PSGY	LG	YQEL	GMGR	HY	PEI	KW	DG	GK	300
Qy	836	OKTSSGA	OL	NEL	V	KYL	LSL	AM	G	H	W	895
Db	301	OKTSSGA	OL	NEL	V	KYL	LSL	AM	G	H	W	360
Qy	896	IHVVAQ	CEEG	ESH	LS	RS	V	K	Y	A	E	955
Db	361	IHVVAQ	CEEG	ESH	LS	RS	V	K	Y	A	E	420
Qy	956	LLKTSW	FF	DV	L	IKS	MAO	HL	I	ENS	KV	1011
Db	421	LLKTSW	FF	DV	L	IKS	MAO	HL	I	ENS	KV	480
Qy	1016	PEASGN	AN	HS	L	AV	F	IK	C	E	T	1077
Db	481	PEASGN	AN	HS	L	AV	F	IK	C	E	T	540
Qy	1076	EHYI	PL	N	P	M	P	E	G	K	R	1133
Db	541	EHYI	PL	N	P	M	P	E	G	K	R	600
Qy	1136	AISV	LN	KL	L	IKS	P	D	R	Y	A	1199
Db	601	AISV	LN	KL	L	IKS	P	D	R	Y	A	660
Qy	1196	VKDE	S	L	A	P	A	V	N	P	L	1255
Db	661	VKDE	S	L	A	P	A	V	N	P	L	720
Qy	1256	LIS	D	S	G	N	S	L	P	R	N	1311
Db	721	LIS	D	S	G	N	S	L	P	R	N	780
Qy	1316	DDAL	P	T	Y	N	K	A	S	T	E	1377
Db	781	DDAL	P	T	Y	N	K	A	S	T	E	840
Qy	1376	NHSY	G	S	D	A	D	V	L	H	Q	1433
Db	841	NHSY	G	S	D	A	D	V	L	H	Q	900
Qy	1436	YL	C	F	I	O	K	H	O	S	E	1499
Db	901	YL	C	F	I	O	K	H	O	S	E	960

Qy	1496	EASQLLVEFLRNNDYDGKSPFRTHLQVLIISVSLIADYVGGIFRFOOSLIINNCCAN	1555
Dp	961	EASQLLVEFLRNNDYDGKSPFRTHLQVLIISVSLIADYVGGIFRFOOSLIINNCCAN	1020
Qy	1556	SDRLIKHTSSSDYKDLTKRIRTVLMAATAQMKHEHNPENLVLDQYSLAASYSTPELRK	1615
Dp	1021	SDRLIKHTSSSDYKDLTKRIRTVLMAATAQMKHEHNPENLVLDQYSLAASYSTPELRK	1080
Qy	1616	TWLDSPARIHVXGDLSEAMCYVHTALVAEYITRKEAVQMBPPLLPHSHSACLRSRG	1675
Dp	1081	TWLDSPARIHVXGDLSEAMCYVHTALVAEYITRKEAVQMBPPLLPHSHSACLRSRG	1140
Qy	1676	GVFPGOGTAFRVITPNPIDEASMMEDVGMQDVFHNEVYLMELLEQCADGLMKAREYELIA	1735
Dp	1141	GVFPGOGTAFRVITPNPIDEASMMEDVGMQDVFHNEVYLMELLEQCADGLMKAREYELIA	1200
Qy	1736	DIYELLIPIYEKRDPERLALHYDTLHRAYSKTYEVMHSGRRLIGTYFRVAFPGQAQOY	1795
Dp	1201	DIYELLIPIYEKRDPERLALHYDTLHRAYSKTYEVMHSGRRLIGTYFRVAFPGQAQOY	1260
Qy	1796	FTDSEIVDEGFPEDEDEKEYITYPEKPTLPSEISQRLKLYSDKFGSENNYHIDSGKN	1855
Dp	1261	FTDSEIVDEGFPEDEDEKEYITYPEKPTLPSEISQRLKLYSDKFGSENNYHIDSGKN	1320
Qy	1856	PKDLDSKYAYIQVTHVLPFDEKELQERKTEPERSHNIIRPFMEPPTQYGRQGGVEEO	1915
Dp	1321	PKDLDSKYAYIQVTHVLPFDEKELQERKTEPERSHNIIRPFMEPPTQYGRQGGVEEO	1380
Qy	1916	CKRRTIITAIHCPEYVKRIIPWYQHTTDLNP1EVALIDENSKYVABLROLCSSAEVDMIK	1975
Dp	1381	CKRRTIITAIHCPEYVKRIIPWYQHTTDLNP1EVALIDENSKYVABLROLCSSAEVDMIK	1440
Qy	1976	LQKLQSSVSVOVNAAGLAAARAFLDNTNKRIPDNVYKLKXEYFRQFVPAACGALAVNE	2035
Dp	1441	LQKLQSSVSVOVNAAGLAAARAFLDNTNKRIPDNVYKLKXEYFRQFVPAACGALAVNE	1500
Qy	2036	RLIKEDQLEYOEEMKANRYREMAKELSEIMHEQI 2068	
Dp	1501	RLIKEDQLEYOEEMKANRYREMAKELSEIMHEQI 1533	
RESULT 9			
US-09-736-960-10			
: Sequence 10. Application US/09736960			
: Patent No. US20020102267A1			
: GENERAL INFORMATION:			
: APPLICANT: Lu, Peter			
: APPLICANT: Garman, Jonathan David			
: APPLICANT: Candia III, Albert Frederick			
: APPLICANT: Arbor Vita Corporation			
: TITLE OF INVENTION: CLASP-5 Transmembrane Protein			
: FILE REFERENCE: 020054-000511US			
: CURRENT APPLICATION NUMBER: US/09/736,960			
: CURRENT FILING DATE: 2001-09-20			
: PRIOR APPLICATION NUMBER: US 60/160,860			
: PRIOR FILING DATE: 1999-10-21			
: PRIOR APPLICATION NUMBER: US 60/162,498			
: PRIOR FILING DATE: 1999-10-29			
: PRIOR APPLICATION NUMBER: US 60/170,453			
: PRIOR FILING DATE: 1999-12-13			
: PRIOR APPLICATION NUMBER: US 60/176,195			
: PRIOR FILING DATE: 2000-01-14			
: PRIOR APPLICATION NUMBER: US 60/182,296			
: PRIOR FILING DATE: 2000-02-14			
: PRIOR APPLICATION NUMBER: US 09/547,276			
: PRIOR FILING DATE: 2000-04-11			
: PRIOR APPLICATION NUMBER: US 60/196,267			
: PRIOR FILING DATE: 2000-04-11			
: PRIOR APPLICATION NUMBER: US 60/196,460			
: PRIOR FILING DATE: 2000-04-11			
: PRIOR APPLICATION NUMBER: US 60/196,527			
: PRIOR FILING DATE: 2000-04-11			

```

/ PRIOR APPLICATION NUMBER: US 60/196,528
/ PRIOR FILING DATE: 2000-04-11
/ PRIOR APPLICATION NUMBER: US 09/687,837
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,503
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,508
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,539
/ PRIOR FILING DATE: 2000-10-13
/ PRIOR APPLICATION NUMBER: US 60/240,543
/ PRIOR FILING DATE: 2000-10-13
/ NUMBER OF SEQ ID NOS: 134
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 10
/ LENGTH: 1534
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: human KIAA1058
US-09-736-960-10

```

```

Query Match      72.9%; Score 7976; DB 10; Length 1534;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1532; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 536 ASGNLDKARFSAIYRQDSNKLSDNDMLKILADFRKREKMAKLPVILGNLDTIDNVSD 595
DB 1 ASGNLDKARFSAIYRQDSNKLSDNDMLKILADFRKREKMAKLPVILGNLDTIDNVSD 60
OY 596 PNYVNSYIPTKQFETCSKTPITFEVEEFVPCIPKHTQPYITYNHLYVYKYLKYSQD 655
DB 61 PNYVNSYIPTKQFETCSKTPITFEVEEFVPCIPKHTQPYITYNHLYVYKYLKYSQD 120
OY 656 KSPAKARNALICIEFDSDSEDSQPLKCIYGRGCVFTRSAPAAVLHNNQPEFDEIK 715
DB 121 KSPAKARNALICIEFDSDSEDSQPLKCIYGRGCVFTRSAPAAVLHNNQPEFDEIK 180
OY 716 IELPTQJHEKHHLITFFPVSCDSSKSGSTKRDVETOVGSWMLPLKDGRTVSEHI 775
DB 181 IELPTQJHEKHHLITFFPVSCDSSKSGSTKRDVETOVGSWMLPLKDGRTVSEHI 240
OY 776 PVSANLPSGYLGYELGMRHVGPEIKWVGKPKLKISTHLVSTVYTDQHLNFFQYC 835
DB 241 PVSANLPSGYLGYELGMRHVGPEIKWVGKPKLKISTHLVSTVYTDQHLNFFQYC 300
OY 836 QRTESGAQALGNELVYKLSLHAMEGHVNIAPLPTILNQLFRVLTAADEAVANVTYRI 895
DB 301 QRTESGAQALGNELVYKLSLHAMEGHVNIAPLPTILNQLFRVLTAADEAVANVTYRI 360
OY 896 IHVVAOCHEEGLSEHLSRYKAYKAPYVASEYKTVHEBELTSMPTTIKPSADFITSK 955
DB 361 IHVVAOCHEEGLSEHLSRYKAYKAPYVASEYKTVHEBELTSMPTTIKPSADFITSK 420
OY 956 LKYSNFFPDVILIKSNAQHLIENSKVKLLRNQFPASYNHAAVETVNMMLPHITQKFRN 1015
DB 421 LKYSNFFPDVILIKSNAQHLIENSKVKLLRNQFPASYNHAAVETVNMMLPHITQKFRN 480
OY 1016 PEASKANNSLAVIFIRCTFPMDRGFVFOQINNYISCFAPGDPKTLFEKFEPLRVCHN 1075
DB 481 PEASKANNSLAVIFIRCTFPMDRGFVFOQINNYISCFAPGDPKTLFEKFEPLRVCHN 540
OY 1076 EHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHFLVGLLREVGTAQOEFREVRIL 1135
DB 541 EHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHFLVGLLREVGTAQOEFREVRIL 600
OY 1136 AISVLKNLLIKSFDSDRYASRSHQARIATLYPLFGLLIENVQRIWADVSPPVNAGMT 1195
DB 601 AISVLKNLLIKSFDSDRYASRSHQARIATLYPLFGLLIENVQRIWADVSPPVNAGMT 660
OY 1196 VVDESLALAVNPLVTPQKSTLDNSLHDLGALISGINSPTTSPININSVANDSRGS 1255
DB 661 VVDESLALAVNPLVTPQKSTLDNSLHDLGALISGINSPTTSPININSVANDSRGS 720

```

```

OY 1256 LISTDSGNSLPERNSEKNSLDKHOSSSTLGSVVRCDKLDQSEIKSLMCFLYILKMS 1315
DB 721 LISTDSGNSLPERNSEKNSLDKHOSSSTLGSVVRCDKLDQSEIKSLMCFLYILKMS 780
OY 1316 DDALFTYNNKASTSELMDFITSEVCLHQFOYMGKRYIARTGMHARLQOLGSLDNLSTF 1375
DB 781 DDALFTYNNKASTSELMDFITSEVCLHQFOYMGKRYIARTGMHARLQOLGSLDNLSTF 840
OY 1376 NNSVGHSDADVHOSLEBANIAVEVCLALDPLSLFTLAFPNQOLADHGNPLMKVADV 1435
DB 841 NNSVGHSDADVHOSLEBANIAVEVCLALDPLSLFTLAFPNQOLADHGNPLMKVADV 900
OY 1436 YLCFLOKHOSETALGNVFTALRSLLIYKPESTFEYEGRADMCALCYEILKCCNSKLSIRT 1495
DB 901 YLCFLOKHOSETALGNVFTALRSLLIYKPESTFEYEGRADMCALCYEILKCCNSKLSIRT 960
OY 1496 EASQLLYFLMRNNDYTGKSFVTRTHLOVIISVSQLIADVIGIGTRFOQSISINNCA 1555
DB 961 EASQLLYFLMRNNDYTGKSFVTRTHLOVIISVSQLIADVIGIGTRFOQSISINNCA 1020
OY 1556 SDRLIKHTSFSSDVKDLTKRIRTVLMATRAOMKEHNDPEMLVDLOYSLAKSYASTPELR 1615
DB 1021 SDRLIKHTSFSSDVKDLTKRIRTVLMATRAOMKEHNDPEMLVDLOYSLAKSYASTPELR 1080
OY 1616 TWLDSMARIHVKNGLSEAMCYVAVTALVAEYLTRKAAVOMEPPLPHSHSACLRARG 1675
DB 1081 TWLDSMARIHVKNGLSEAMCYVAVTALVAEYLTRKAAVOMEPPLPHSHSACLRARG 1140
OY 1676 GVFRQCTAFVITPNIDEASMMEDVGMQVHNEVDVLMELBOCADGLMAERYELIA 1735
DB 1141 GVFRQCTAFVITPNIDEASMMEDVGMQVHNEVDVLMELBOCADGLMAERYELIA 1200
OY 1736 DIYKLIPIYERKRPFERLAHYDTLHRAYSKVTEVMSGRLIGTYRVAFFGQAQYO 1795
DB 1201 DIYKLIPIYERKRPFERLAHYDTLHRAYSKVTEVMSGRLIGTYRVAFFGQAQYO 1260
OY 1796 FTDSSETDVEGFEDDEKKEYIYKEBKLPPLSEISORLKLKLYSDKGSNNVMKI QDSGKVN 1855
DB 1261 FTDSSETDVEGFEDDEKKEYIYKEBKLPPLSEISORLKLKLYSDKGSNNVMKI QDSGKVN 1320
OY 1856 PKDLDSKAYIQTHTVIPPPEDEKLOERKTEPERSHNIIRFMEMPFTQTKRGQGVBEQ 1915
DB 1321 PKDLDSKAYIQTHTVIPPPEDEKLOERKTEPERSHNIIRFMEMPFTQTKRGQGVBEQ 1380
OY 1916 CKRRTILTAHCFPVYKRIIPVMVOHHTDMLPIEVAIDEMSKVAELROLCSAENVMIK 1975
DB 1381 CKRRTILTAHCFPVYKRIIPVMVOHHTDMLPIEVAIDEMSKVAELROLCSAENVMIK 1440
OY 1976 LQKLQGSVSQVNVNAGPLAVARAFLDNTNTRKYPDNKYKLYKEVFRQFVEACQALAVNE 2035
DB 1441 LQKLQGSVSQVNVNAGPLAVARAFLDNTNTRKYPDNKYKLYKEVFRQFVEACQALAVNE 1500
OY 2036 RLKEDOLEYOEMKANYREMAKELSEIMHEOI 2068
DB 1501 RLKEDOLEYOEMKANYREMAKELSEIMHEOI 1533

```

```

RESULT 10
US-09-736-968A-9
/ Sequence 9, Application US/09736968A
/ Patent No. US20020169283A1
/ GENERAL INFORMATION:
/ APPLICANT: Lu, Peter
/ APPLICANT: Garman, Jonathan David
/ APPLICANT: Candia, Ili, Albert Frederick
/ APPLICANT: Arbor Vita Corporation
/ TITLE OF INVENTION: CLASP-7 Transmembrane Protein
/ FILE REFERENCE: 020054-000611US
/ CURRENT FILING DATE: US/09/736,968A
/ CURRENT APPLICATION NUMBER: 1255
/ PRIOR APPLICATION NUMBER: US 60/160,860
/ PRIOR FILING DATE: 1999-10-21

```


; PRIOR APPLICATION NUMBER: US 60/162,498
 ; PRIOR FILING DATE: 1999-10-29
 ; PRIOR APPLICATION NUMBER: US 60/170,453
 ; PRIOR FILING DATE: 1999-12-13
 ; PRIOR APPLICATION NUMBER: US 60/176,195
 ; PRIOR FILING DATE: 2000-01-14
 ; PRIOR APPLICATION NUMBER: US 60/182,296
 ; PRIOR FILING DATE: 2000-02-14
 ; PRIOR APPLICATION NUMBER: US 09/547,276
 ; PRIOR FILING DATE: 2000-04-11
 ; PRIOR APPLICATION NUMBER: US 60/196,267
 ; PRIOR FILING DATE: 2000-04-11
 ; PRIOR APPLICATION NUMBER: US 60/196,460
 ; PRIOR FILING DATE: 2000-04-11
 ; PRIOR APPLICATION NUMBER: US 60/196,527
 ; PRIOR FILING DATE: 2000-04-11
 ; PRIOR APPLICATION NUMBER: US 60/196,528
 ; PRIOR FILING DATE: 2000-04-11
 ; PRIOR APPLICATION NUMBER: US 09/687,837
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: US 60/240,503
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: US 60/240,508
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: US 60/240,539
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: US 60/240,543
 ; PRIOR FILING DATE: 2000-10-13
 ; NUMBER OF SEQ ID NOS: 115
 ; SOFTWARE: Patent Ver. 2.1
 ; SEQ ID NO: 9
 ; LENGTH: 1352
 ; TYPE: PRF
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: human CLASP-2A
 ; US-09-736-968A-9

Query Match 62.0%; Score 6784.5; DB 9; Length 1352;
 Best Local Similarity 92.7%; Pred. No. 0;
 Matches 1355; Conservative 1; Mismatches 3; Indels 101; Gaps 3;

QY 701 VLAHHNPEFYDEIKIELPTOLHEKHLITTFPHVSCDSKSGSTKRDVETOVGYSWL 760
 DB 1 VLAHHNPEFYDEIKIELPTOLHEKHLITTFPHVSCDSKSGSTKRDVETOVGYSWL 60
 QY 761 PLKGGRRVTSSECHIPVSANLPSGUYGYOELGNGRHYGEITKWDGKFLKISTHLVST 820
 DB 61 PLKGGRRVTSSECHIPVSANLPSGUYGYOELGNGRHYGEITKWDGKFLKISTHLVST 120
 QY 821 VTTODHNLNPFYOCCOKTSSGAOLGNELVKYLSLHAMEGHVMTAFLPTIINOLFVRLT 880
 DB 121 VTTODHNLNPFYOCCOKTSSGAOLGNELVKYLSLHAMEGHVMTAFLPTIINOLFVRLT 180
 QY 881 RATOEEVAVNTRVLIHVVAQCHBEGLESLSYVYKA/KASPYVASEYKTVHEBELTKSM 940
 DB 181 RATOEEVAVNTRVLIHVVAQCHBEGLESLSYVYKA/KASPYVASEYKTVHEBELTKSM 240
 QY 941 TTIILKPSADPLTSNKLKTSWFFPDVLTISMAOHLIENSKYKLLRNQRPASTYHAAETV 1000
 DB 241 TTIILKPSADPLTSNKLKTSWFFPDVLTISMAOHLIENSKYKLLRNQRPASTYHAAETV 300
 QY 1001 VMLMLPHITQKFRDNPESKNNHSLAVFIKRCFTMDRGPFVKOINNYISCAPDDEPT 1060
 DB 301 VMLMLPHITQKFRDNPESKNNHSLAVFIKRCFTMDRGPFVKOINNYISCAPDDEPT 360
 QY 1061 LEEYKEFLRVLVVCNHEHYIPLNLPMPFGKRIQRYODLQLDVSLTDEFCRNHFLVGLLIR 1120
 DB 361 LEEYKEFLRVLVVCNHEHYIPLNLPMPFGKRIQRYODLQLDVSLTDEFCRNHFLVGLLIR 420
 QY 1121 EVGTALQERREVRLLAISVLKULLIGASDDRYASASHQARIATVLYPLFGILLIENVQRI 1180
 DB 421 EVGTALQERREVRLLAISVLKULLIGASDDRYASASHQARIATVLYPLFGILLIENVQRI 480

QY 1181 NVRDVSPPPVNAGMTVKDESIALPAVNPLVTPQKSTLDNSLHKDLKALISGIASPYTTS 1240
 DB 481 NVRDVSPPPVNAGMTVKDESIALPAVNPLVTPQKSTLDNSLHKDLKALISGIASPYTTS 540
 QY 1241 TPININSVANADSRGSLISTDSGNSLPERNSEKNSLDHQOOSTLGNSVVRCDKLDQSEI 1300
 DB 541 TPININSVANADSRGSLISTDSGNSLPERNSEKNSLDHQOOSTLGNSVVRCDKLDQSEI 600
 QY 1301 KSLMCFIYIILKSNBDDLFYWNKASTSELMDFPTISEVCLHQOYKGRYIAQNOGEL 1354
 DB 601 KSLMCFIYIILKSNBDDLFYWNKASTSELMDFPTISEVCLHQOYKGRYIAQNOGEL 660
 QY 1355 -----RTGMMHARLQOLGSLDNLSTFNHSGHSDADVLHQSLLEANIA 1397
 DB 661 GPVYHDKRSQTLPVGRNRTGMHARLQOLGSLDNLSTFNHSGHSDADVLHQSLLEANIA 720
 QY 1398 TEVCLTALDLSLFTLAFKNOLLADHGNPLMKVFDVYLCPLOKHGSETALKONVFTALR 1457
 DB 721 TEVCLTALDLSLFTLAFKNOLLADHGNPLMKVFDVYLCPLOKHGSETALKONVFTALR 780
 QY 1458 SLIYKPSSTFYEGRADMCALCYELKCCNSKLSIRTEASQOLYFLARNNPDYTGKXSF 1517
 DB 781 SLIYKPSSTFYEGRADMCALCYELKCCNSKLSIRTEASQOLYFLARNNPDYTGKXSF 840
 QY 1518 VRTHLQVITISVSQLIADVVGIGTRFOOSLSTINNCASDRLIKRTSPSSDVKDLTKRIR 1577
 DB 841 VRTHLQVITISVSQLIADVVGIGTRFOOSLSTINNCASDRLIKRTSPSSDVKDLTKRIR 900
 QY 1578 TVLMATAQMKHEHNDPEMLVDLQYSIAKSYASTPELRKTMVLSMARIHVKNQDLSBAAMC 1637
 DB 901 TVLMATAQMKHEHNDPEMLVDLQYSIAKSYASTPELRKTMVLSMARIHVKNQDLSBAAMC 960
 QY 1638 YVHTALVAEYLITREANQVMBPRLPHSHSACLRSGVRCQGTARVITPNIDENAS 1697
 DB 961 YVHTALVAEYLITREANQVMBPRLPHSHSACLRSGVRCQGTARVITPNIDENAS 997
 QY 1698 NMEDVGMQDVHFNEDVLMELLBOCADGMLKABRYELADIYKLIPIYEKRDQFRLAHL 1757
 DB 998 NMEDVGMQDVHFNEDVLMELLBOCADGMLKABRYELADIYKLIPIYEKRDQFRLAHL 1051
 QY 1758 YDTHRAYSKYTEVNHSGRLLGTFRVAFGQAAYOFTDSETDVEGFEDDEGKEYIY 1817
 DB 1052 -----FEDEGKEYIY 1062
 QY 1818 KEPKLTPLSEISORLLKYSDFGSENYKMTODSKNPKULDSKATYIOTHTVPPFDE 1877
 DB 1063 KEPKLTPLSEISORLLKYSDFGSENYKMTODSKNPKULDSKATYIOTHTVPPFDE 1122
 QY 1878 KELDERKTEPFRSHNIRBFEMPTOTGKRGOGGVEGCKRRTITLAIHCPYVKGRIY 1937
 DB 1123 KELDERKTEPFRSHNIRBFEMPTOTGKRGOGGVEGCKRRTITLAIHCPYVKGRIY 1182
 QY 1938 MYOHTDNLPIEVALIDENSKVABLRQCSSAEVDMILQULQGSVSVQVNAAGLAYAR 1997
 DB 1183 MYOHTDNLPIEVALIDENSKVABLRQCSSAEVDMILQULQGSVSVQVNAAGLAYAR 1242
 QY 1998 AFLDPTNTKRYPDNKVKLLKXVFPFVACGQALAVNERLIKEDOLEYQEBKANYREMA 2057
 DB 1243 AFLDPTNTKRYPDNKVKLLKXVFPFVACGQALAVNERLIKEDOLEYQEBKANYREMA 1302
 QY 2058 KELSEIMHEOICPLEEKTSVLPNSLHIFNATISGPTSTMVGMNTSSSSSV 2107
 DB 1303 KELSEIMHEOICPLEEKTSVLPNSLHIFNATISGPTSTMVGMNTSSSSSV 1352

RESULT 11
 US-09-736-969A-9
 ; Sequence 9, Application US/09736969A
 ; Patent No. US20020068302A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lu, Peter
 ; APPLICANT: Garman, Jonathan David

APPLICANT: Candia III, Albert Frederick
APPLICANT: Arbor Vita Corporation
TITLE OF INVENTION: CLASP-4 Transmembrane Protein
FILE REFERENCE: 020054-00041US
CURRENT FILING DATE: 2000-12-13
CURRENT FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 60/160,860
PRIOR FILING DATE: 1999-10-21
PRIOR APPLICATION NUMBER: US 60/162,498
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: US 60/170,453
PRIOR FILING DATE: 1999-12-13
PRIOR APPLICATION NUMBER: US 60/176,195
PRIOR FILING DATE: 2000-01-14
PRIOR APPLICATION NUMBER: US 60/182,296
PRIOR FILING DATE: 2000-02-14
PRIOR APPLICATION NUMBER: US 09/547,276
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,267
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,460
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,527
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 60/196,528
PRIOR FILING DATE: 2000-04-11
PRIOR APPLICATION NUMBER: US 09/687,837
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,503
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,508
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,539
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,543
PRIOR FILING DATE: 2000-10-13
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 9
LENGTH: 1352
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: human CLASP-2
US-09-736-969A-9

Query Match 62.0%; Score 6784.5; DB 10; Length 1352;
Best Local Similarity 92.7%; Pred. No. 0;
Matches 135; Conservative 1; Mismatches 3; Indels 101; Gaps 3;
Db 701 VHHHNPPEFYDIKIELPTQLHEKHLITLFFHVS CDNSKSGSTKKRUVETQVGSWL 760
1 VHHHNPPEFYDIKIELPTQLHEKHLITLFFHVS CDNSKSGSTKKRUVETQVGSWL 60
Qy 761 PLKDGRRVTSSEGIIVSANLPSGYLGYOELGWRHYGPEIKWVDGKPLKISTHLVST 820
61 PLKDGRRVTSSEGIIVSANLPSGYLGYOELGWRHYGPEIKWVDGKPLKISTHLVST 120
Db 821 VTTODHNLNFPYCKCTESSGAOLGNELVKTKSLHAMEGHMIAFLPTIINQLFRVLT 880
121 VTTODHNLNFPYCKCTESSGAOLGNELVKTKSLHAMEGHMIAFLPTIINQLFRVLT 180
Qy 881 RATOEEVAVNVRVLIHHVAAQCHEBGLBSHLSSYVAYKAPRYA SEKYTVHEBLTSM 940
181 RATOEEVAVNVRVLIHHVAAQCHEBGLBSHLSSYVAYKAPRYA SEKYTVHEBLTSM 240
Db 941 TTIILKSADELTSNKLKTSWFFDVLIKSMQHLIENSKVKLLRNQRPASYHAAVETV 1000
241 TTIILKSADELTSNKLKTSWFFDVLIKSMQHLIENSKVKLLRNQRPASYHAAVETV 300
Qy 1001 VNMMLPHITQKFRDNEASKNANHSIAVTKRCFTTMDRGFVFKQINNYISCFAPDDPT 1060
301 VNMMLPHITQKFRDNEASKNANHSIAVTKRCFTTMDRGFVFKQINNYISCFAPDDPT 360

Qy 1061 LFEYKFEFLRVYCNHEHYIPLNLPMFPFGKRIQRYQDQLDYSLTDECRNHFVGLLLR 1120
Db 361 LFEYKFEFLRVYCNHEHYIPLNLPMFPFGKRIQRYQDQLDYSLTDECRNHFVGLLLR 420
Qy 1121 EVGTALQEFREVRLLAISVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVORI 1180
Db 421 EVGTALQEFREVRLLAISVLKNLLIKHSFDDRYASRSHQARIATLYLPLFGLLIENVORI 480
Qy 1181 NRVDSPPPVNAGMVKVCKESLALPAVNPLVPQKSTLDNSLHMDLGAISGASPYTS 1240
Db 481 NRVDSPPPVNAGMVKVCKESLALPAVNPLVPQKSTLDNSLHMDLGAISGASPYTS 540
Qy 1241 TPINISVNNADSRGSLISTDSGNSLPERNSEKNSLDRHQOOSTLGNSVVRCDKLDQSEI 1300
Db 541 TPINISVNNADSRGSLISTDSGNSLPERNSEKNSLDRHQOOSTLGNSVVRCDKLDQSEI 600
Qy 1301 KSLMCFYIILKSMDDALFTYWNKASTSEIMDFTTISEVCLHOFQYWGKRYIA----- 1354
Db 601 KSLMCFYIILKSMDDALFTYWNKASTSEIMDFTTISEVCLHOFQYWGKRYIARNOGL 660
Qy 1355 -----RTGMHARLQOLGSLDNSLTNHSYGSDADVLHQSLEANIA 1397
Db 661 GPIVHDRKSQTLPVSRNRTGMHARLQOLGSLDNSLTNHSYGSDADVLHQSLEANIA 720
Qy 1398 TEVCUTLADTLSLFTLAFKQQLADHGNPLMKVFVDYLCFLQKHQSETALKVFTALR 1457
Db 721 TEVCUTLADTLSLFTLAFKQQLADHGNPLMKVFVDYLCFLQKHQSETALKVFTALR 780
Qy 1458 SLIYFPSTFYEGRADMCALCYEILKCCNSKLSIRTEASQQLYFLMRNPDYTGKXSF 1517
Db 781 SLIYFPSTFYEGRADMCALCYEILKCCNSKLSIRTEASQQLYFLMRNPDYTGKXSF 840
Qy 1518 VPTHQVILISVQLADVVGIGTFPOOSLIINNCAUSDLLIKHTSSSDVKDLTKRIR 1577
Db 841 VPTHQVILISVQLADVVGIGTFPOOSLIINNCAUSDLLIKHTSSSDVKDLTKRIR 900
Qy 1578 TVLMATAQMKHEHNDPEMLVDLQYSLAKSYASTPELRRTWLDMSMARIHVKNGDLSSEAAAC 1637
Db 901 TVLMATAQMKHEHNDPEMLVDLQYSLAKSYASTPELRRTWLDMSMARIHVKNGDLSSEAAAC 960
Qy 1638 YVHTALVAEYLITREKAVQWEPPLPHSHSACLRSSRGVFRQGTAFRVITPNIDEAS 1697
Db 961 YVHTALVAEYLITREKAVQWEPPLPHSHSACLRSSRGVFRQGTAFRVITPNIDEAS 997
Qy 1698 NMEDVGMQDVHFNEDVLMELLEQCADGLMKAREYELADIYKLIIPYEKRDPERLNL 1757
Db 998 NMEDVGMQDVHFNEDVLMELLEQCADGLMKAREYELADIYKLIIPYEKRDPERLNL 1051
Qy 1758 YDTLHRAYSKTVEMHSGRRLIGTYFRVAFGQAQYQFTDSEITDVEGFFDEDEKEXIY 1817
Db 1052 -----FDEDEKEXIY 1062
Qy 1818 KEPKLTPLSEISQRLIKLKYSPKSGSENVKMIQDSGRVNPCKLDISKYAVIQTHTVIPFDE 1877
Db 1063 KEPKLTPLSEISQRLIKLKYSPKSGSENVKMIQDSGRVNPCKLDISKYAVIQTHTVIPFDE 1122
Qy 1878 KELQKRTKPEFSRSHIRRFEMFPTQNGKQOGGVEOCCRRITLTALHCPFYVKRIRPV 1937
Db 1123 KELQKRTKPEFSRSHIRRFEMFPTQNGKQOGGVEOCCRRITLTALHCPFYVKRIRPV 1182
Qy 1938 MYOHTDNLNPIEVALIDEMSKYAEELQCSSAEVDMIKLQKLOGSVSVQVNAAGLAAAR 1997
Db 1183 MYOHTDNLNPIEVALIDEMSKYAEELQCSSAEVDMIKLQKLOGSVSVQVNAAGLAAAR 1242
Qy 1998 AFLDNTNTRYPPDNKVKLLKEVFRQFVACGQALAVNRLIKEDQLEYOEEMKANRYREMA 2057
Db 1243 AFLDNTNTRYPPDNKVKLLKEVFRQFVACGQALAVNRLIKEDQLEYOEEMKANRYREMA 1302
Qy 2058 KELSIMEOICPLEBKSYSVPLNSLHIFNAISGPTSTNMHGMSSSSSV 2107
Db 1303 KELSIMEOICPLEBKSYSVPLNSLHIFNAISGPTSTNMHGMSSSSSV 1352

QY 1998 AFLDDNTTKRYYPNDKVKLLKEVFRQFVEACGQALAVNERLIKEDOLEYOBEMKANRYRENA 2057
 DB 1243 AFLDDNTTKRYYPNDKVKLLKEVFRQFVEACGQALAVNERLIKEDOLEYOBEMKANRYRENA 1302
 QY 2058 KEISEIMHEOICPLEKTSYLPNSLHIFNIAISTPTSTVHGHTSSSSV 2107
 DB 1303 KEISEIMHEOICPLEKTSYLPNSLHIFNIAISTPTSTVHGHTSSSSV 1352

RESULT 13
 US-09-978-244A-10
 ; Sequence 10, Application US/09978244A
 ; Publication No. US20030103992A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Lu, Peter S
 ; APPLICANT: Garmen, Jonathan D.
 ; APPLICANT: Candia Jii, Albert F.
 ; APPLICANT: Arbor Vita Corporation
 ; TITLE OF INVENTION: CLASP MEMBRANE PROTEINS
 ; FILE REFERENCE: 020554-000161US
 ; CURRENT APPLICATION NUMBER: US/09/978,244A
 ; PRIOR FILING DATE: 2001-10-15
 ; PRIOR APPLICATION NUMBER: US 60/310,028
 ; PRIOR FILING DATE: 2001-08-03
 ; PRIOR APPLICATION NUMBER: US 09/737,246
 ; PRIOR FILING DATE: 2000-12-13
 ; PRIOR APPLICATION NUMBER: US 09/736,969
 ; PRIOR FILING DATE: 2000-12-13
 ; PRIOR APPLICATION NUMBER: US 09/736,960
 ; PRIOR FILING DATE: 2000-12-13
 ; PRIOR APPLICATION NUMBER: US 09/736,968
 ; PRIOR FILING DATE: 2000-12-13
 ; PRIOR APPLICATION NUMBER: US 60/240,545
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: US 60/240,508
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: US 60/240,503
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: US 60/240,539
 ; PRIOR FILING DATE: 2000-10-13
 ; PRIOR APPLICATION NUMBER: US 60/240,543
 ; PRIOR FILING DATE: 2000-10-13
 ; Remaining Prior Application data removed - See file wrapper or PAM.
 ; NUMBER OF SEQ ID NOS: 106
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 10
 ; LENGTH: 2073
 ; TYPE: PRT
 ; ORGANISM: Mus sp.
 ; FEATURE:
 ; OTHER INFORMATION: CLASP-4
 US-09-978-244A-10

Query Match 58.3%; Score 6375.5; DB 9; Length 2073;
 Best Local Similarity 58.8%; Pred. No. 0;
 Matches 1257; Conservative 325; Mismatches 425; Indels 131; Gaps 28;

QY 11 AETRFKTRALSXPGTAELROSYSSEVVRGSVYLAKLLEPLDYENVVIOKQTQINDL 70
 DB 2 AETRFKTRALSXPGTAELROSYSSEVVRGSVYLAKLLEPLDYENVVIOKQTQINDL 61
 QY 71 REMLLFPYDDPOTALLRQGRYICSTVPAKAEENASLFTVTECITKYNSDMLVNYKYD 130
 DB 62 RDLMPMPMDISIVIGRQRTVOSTVPEDAEKRAQSLFYKECITKYSTDMHVVNYKYD 121
 QY 131 YGGEFQQLFNKVKLKLKLVHYVEVDEVDKDDAASLCSQKGTGKHGMVLYKGNNSAI 190
 DB 122 FSGDFPMLPCKSKRPKIPBNHVEIDEDEKEDSSSLCSQKGVYKQGMVLYKGNNSAI 181
 QY 191 SYTMRSFKRFPFLIOLGSGSYKFEFLKLO--KEKGSIFLGLYGVYFRNNKVRRAF 248
 DB 182 TYTMKFKRRFYLTQLPGSYYILNSYKQKSKSKGCIYLDACIDV--VQCFKMRRAAF 240

QY 249 ELKMODKSSYLLAADSEVEMEMETITLNLKIQLNFEAAMOERKNG---DSHEDDEQSL 304
 DB 241 ELKMDLKHSHYLAABTEDEMMEMLMLKKITQLINDSLVQEKQOVVEALIQEETSQOKA 300
 QY 305 EGSGGGLD-SYLPBLAKAREAE--IKL-KSESRVLYLDPDAOKLDPSSAEPVKSFE 360
 DB 301 ENIMASLERSMHELMKRGRETEQLNKLRSRGGRGNLPSFEVQRLPFGSEIDPVKFE 360
 QY 361 EKGKRIIVKQNDLSFNLQCCVAENEBGPTNVPEFVTLSPFDIKYNRKISADPHVTLN 420
 DB 361 EKCNRKRFVNCCHDLFTNLIGDANKGPTVNEPFTNLALFDYKNNCKISADPHVDLN 420
 QY 421 HFSVRQMIATTSPALMN-----GSGPETOSALRGILHEAAMQPRQGSFVYTCPPDIFL 475
 DB 421 PPSVAREMLMGSTQSLNNGNAKGFSPE--SLIHGIA-ESQLCYIKQGISVYNNPHEIFL 477
 QY 476 VARIKVTLOGSTTHCAEPYMKSSDSKVAQKYLKAKACQRLGQYRMPFMAAFTLEKD 535
 DB 478 VRIKVKVLQGNITHCAEPYIKNSDPIKTAQKVHRTAKQVCSRLGQYRMPFMAAARIFPD 537
 QY 536 ASGNLDKNAFSAIYRQDSNKLSDNDMLKLADPKRPEKMAKLPYLGNLDTIDNVSSD 595
 DB 538 VQGSIDLGRFSPLKQDSSKLSNEDILKLSEYKPEK-TKLQIIPQOLSTIVCEVPYD 596
 QY 536 FBNVNSSYIPTKQPE-TCSKTPITFEVEEFVPCIPKHTOPYTYTNHLYVYPKYLYDS 654
 DB 597 LENCITSSGYPLKPEKNCQN--IIVEVEEFVPEMTKCYPTTIKNNHLYVYPLQKXDS 654
 QY 655 QKSFKAARNIAICIFKQSDSEDSQPLKCIYGRPGRPFTSAPAAVLHHNNEFFYDEI 714
 DB 655 QKSFKAARNIACVCFRSDSDSALKCIYKPAVSFTTNAAVAVSHNQNPEFYDEI 714
 QY 715 KIELPTOLHEKNNLLTFEFHNSCDNSKSGSTKRPVVEFOYGSYPLKDKGRVYTSSEH 774
 DB 715 KIELPIHLHOKNNLLFTYHNSCEINTGKTKQDTVTEPVGFAVPLKDKGRVLTLEQ 774
 QY 775 IPVSANLPSGYLGYOELGKGRHYGPEIKMVDSGKELKISTHLYSTVYVTDQHLNPFQY 834
 DB 775 LPVSANLPGVLYNVNDASRQSNADIKMVGAKELKIKTHLESTIYTQDLHVAKFFHH 834
 QY 835 CQKTESGADALGNELVKTLSLHAMEGVMTAFELTTLNOLFRVLTATQF-EVAVNATR 893
 DB 835 COLIOGSGKEVGEELIKYKLCIHAMEIQVMOFLVLIIMOLFVLTNNTHBEDVINCIM 894
 QY 894 VTIHVACQHEGESHLSRYVYKAYKAEPVYASRYKTNHEELISMTTLIKPSDFLTS 953
 DB 895 VLIHIVSKQHEGELSYLSRFSIKSFPEKSTLQDLNHELTATTMALIKQSDFLAI 954
 QY 954 NKLTKYSWFFPDVLIKSMAOHLIENSKYVLLRNORFPASYNHNAVETVNMMLPHITOKR 1013
 DB 955 NKLTKYSWFFPELIKSMATYLLLENKIKLPRGQFFPEAHNVHLSLLAIIPHYTIRYA 1014
 QY 1014 DNEPASKANNSLAVITRGTFTMDRGVFKOINNYISCFRABDKTLPYKFEFLAYVC 1073
 DB 1015 EIPDESRRGNYSLSLFLKRCITLMDRGVFLINDYISGFSPKBPVLAKEYFEFLQIC 1074
 QY 1074 NNEHYTILNLPMPPEKGRIOYODLQYSLTDECRNHPVYGLLRVETGALREFRFR 1133
 DB 1075 NNEHYTILNLPMPAPKPKQRYQVDSNLEYSLSDEKCNHFLVGLLRKTSIALQDNYEIR 1134
 QY 1134 LIAISVLNKLIIKISFDDRYASRSHOARIATLYLPFLGLLIENVQIRNVRD---SPFP 1189
 DB 1135 YTAISVINKLLIKAFDRYQHNQQAIAQIOLYLPFVGLLENNIQLRLAGRDTLYSCAMP 1194
 QY 1190 VNAGCTVQDESIALPA--VNPVLTVPQKSTLDNSLHKOLLGALISGIASPYTT--STPNI 1244
 DB 1195 SSAS---SDE--FPCGFSVP---TNRGS-----LASDKOTAVGSFONG 1229
 QY 1245 NSVRNADRSGLISTDSGNSLPERNSEKNSGLDKHQSGSTLGNSSVVRCDKLDQSEIKSL 1304
 DB 1230 HGIKREDSRGSILT-DEGATGFPDPGSENT---ROSSS-RSSVSQYNRDLQYETIRNLL 1283
 QY 1305 MCFYILTKMSDDALFTYMNKASTELMDFTLISEVCLHOFQYMGKRYIA----- 1354

```

Db 1284 MCYLIVKVISSEDTLLTYWNKVSPOELINILVLEVCLEPHFRYMGKRNIFARVHDAWLSKH 1343
Qy 1355 -----RTGMHARLOOLGSLDNLFNHSHY3SDADVLSHSLNANATE 1399
Db 1344 FGDIDKSCQMPALRRRSVGMARLOHLSLESSFTLNHSSATTEDIFQALLBEGTATE 1403
Qy 1400 VCLTALDLTSLFTLAKQQLADHGNPLMKYFVYVLCFLQKHOSETALKNVFTALNSL 1459
Db 1404 VSLTLETISFTQCFKQQLNNDGNHPLMKVFDIHLAFLANGOSEVSLKIVPALSRSF 1463
Qy 1460 IYKPFSTFEGRADMCALCTELKCNKSLSTRTSASQLLYFLMRNFDYTGKKSFR 1519
Db 1464 ISKPSAFKFGKVNCAFCYEVLCCKTSKISSTRMSALYLLMRNFEYTKRTFLR 1523
Qy 1520 THLOVYISVSLIADVIGIGTRFOOSLIINNANSDDLHHTSSPSDYKLTIRIV 1579
Db 1544 THLOVYISVSLIADVIGIGTRFOOSLIINNANSDFHKAIAFPIEVKLTIRIV 1583
Qy 1580 LMATAQMKHENDPEMLVDLOYSLAKSYASTPELRKTMWLSMARLHVKNGLSEAMCY 1639
Db 1584 LMATAQMKHENDPEMLVDLOYSLAKSYASTPELRKTMWLSMAKIHKNGLSEAMCY 1643
Qy 1640 HTTALVAETLTRKEAVQEPRLPHSHSACLRSSGCVFRQCTAFRVTENIDEBASM 1699
Db 1644 HVAALVAEFLHRKK-----LFPSCSAFKITPNIDEGAMK 1680
Qy 1700 EDVGNQDVHFNEDVLMELLEGADGIMKERYELADIYGLIPIYERKRDPERLALYD 1759
Db 1681 EDAGMDVAYSEVLELEQCVDGIMKERYEVSEISGLIPIYERKREKLTQYR 1740
Qy 1760 TLHRAYSKTEVMSHSGRLLGTYFRVAFGQAQOFTD3ETDVGSGFDEDEGKEYIYE 1819
Db 1741 TLHGAVTKLEVMHTKKRLGTFRVAFFGQS-----FFDEEGKEYIYE 1786
Qy 1820 PCLTULSEISQRLKLYSDKFSSENVKMIQDSGKVNPKDLSKAVIQTVAIIPFDEKE 1879
Db 1787 PCLTULSEISQRLKLYSDKFSSENVKMIQDSGKVNPKDLSKAVIQTVAIIPFDEKE 1846
Qy 1880 LOERTPEFRSHNIRFMEPMTOTGRKQSGVEBOCKRRTILTAHCPYKRIIPVY 1939
Db 1847 LHERKTEBERNNINRFPVEADYTLISGKQGLCEOCCKRRTILTTSSNFPYKRIIPVY 1906
Qy 1940 OHHTDLNPLEVALIDEMSKVAELRQLCSSAEYDMIKLQKLOGSVSVQNAAGPLAVARAF 1999
Db 1907 EGVVNLKPIDVATDEIKCKTALHLKCSSVDVDMIQLOKLOGSVSVQNAAGPLAVARAF 1966
Qy 2000 LDDTNTKRYPDNKKVLLKEVFRQVYACGQALAVNERLIKEDOLEYQEBMKANYREMAKE 2059
Db 1967 LNESQANKYPRPKVNELKQMFKFIQACISALELNERLIKEDOLEYHEGLKSNPRDMYKE 2026
Qy 2060 LSEIHQEOCLPEKTSVLP---NSLHTNALSIPPTS 2094
Db 2027 LSDIHEQI--LQEDTMHSPMNNLTLVHCAISGTSN 2062

```

```

RESULT 14
US-09-736-968A-105
; Sequence 105, Application US/09736968A
; Patent No. US20020169283A1
; GENERAL INFORMATION:
; APPLICANT: Lu, Peter
; APPLICANT: Garman, Jonathan David
; APPLICANT: Candia III, Albert Frederick
; APPLICANT: Arbor Vita Corporation
; TITLE OF INVENTION: CLASP-7 Transmembrane Protein
; FILE REFERENCE: 020054-000611US
; CURRENT APPLICATION NUMBER: US/09/736,968A
; CURRENT FILING DATE: 2000-12-13
; PRIOR APPLICATION NUMBER: US 60/160,860
; PRIOR FILING DATE: 1999-10-21
; PRIOR APPLICATION NUMBER: US 60/162,498
; PRIOR FILING DATE: 1999-10-29

```

```

; PRIOR APPLICATION NUMBER: US 60/170,453
; PRIOR FILING DATE: 1999-12-13
; PRIOR APPLICATION NUMBER: US 60/176,195
; PRIOR FILING DATE: 2000-01-14
; PRIOR APPLICATION NUMBER: US 60/182,296
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: US 09/547,276
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,267
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,460
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,527
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 60/196,528
; PRIOR FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 09/687,837
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,503
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,508
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,539
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: US 60/240,543
; PRIOR FILING DATE: 2000-10-13
; NUMBER OF SEQ ID NOS: 115
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 105
; LENGTH: 2008
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: human CLASP-4
US-09-736-968A-105

Query Match 55.8%; Score 6105.5; DB 9; Length 2008;
Best Local Similarity 57.8%; Pred. No. 0;
Matches 1197; Conservative 327; Mismatches 415; Indels 133; Gaps 24;

Qy 75 LPPDDFOTALLRQGRIGSTVPAKAEERQSLFTECITTSNDMLVNYKEDYGE 134
Db 1 MPEMEDISVIGRRRTVOSTVEDAKRAQSLFVKECITYSDMHWVYKXEDFSGD 60
Qy 135 FROLNKKVKKDKLPVHYVEDEVDKDEDAASIGSQSGITKGMVYKGMNNSAISTYM 194
Db 61 FRLPCKSLRPEKLPNHVFEIDEDCKEDSSLSQSGVYKQGMVYKGMNNSAISTYM 120
Qy 195 RSFKRPFHLLQLDGSGYKFEFLKDLQ--KEPKSIFLGPLYGVSRNNKYRFAFELKM 252
Db 121 KVFRRRYVYVQLQDPGSYILNSYKDEKNSKSCICYIDACIDV--VQPKRRHAFELKM 179
Qy 253 QDKSSYLLAADSEVMEEMWITLNLKIQLNFEAAMQEKRNG--DSHEDD---QSTLGGSG 308
Db 180 LDKYSHYLAETEQEMEEWILTLLKTIQINDSLVQEKKTVEYTAQDDETSQGAENIM 239
Qy 309 SGLD-STYPELAKARAE--IKL-KSESRYKLFYLDPADKLDPSASAPRYKSEERKG 364
Db 240 ASLERSMHPMLMKYGRTEQUNLKSIRDGRQNLFSFSDVGRLPDPSGIBPDIKPEEKN 299
Qy 365 KRILVKNDSLNFNLQCCVAENEBGPTTVVPEFFVTLSLFDIKYRNKISADPFVDLNFHSV 424
Db 300 KRLVNCCHDLNLFNLIGQGDNAKGRPTNVPEFFINLAFVKNCKISADPFVDLNPFSV 359
Qy 425 RQMTATSPALMN-----GSGPETQSALRGILHEAAMYPRQGISVTCPPHDFLVARI 479
Db 360 REMLWGSSTQLASGSPKGSPE--SYHNGIA--ESQLRAYIQGGIFSVTPNPBDFLVARI 416
Qy 480 EKVLAGSTTHCAEPYMSSDSSKVAQKVLKNAKQACQRLGQYRMPAARLPFADAGN 539
Db 417 EKVLAGSTTHCAEPYKNSDPVKTQKVRKAKOVCSLGGYRMPAARLPFADAGN 476
Qy 540 LDKNARFSAIYRQDSNKLSDNDMLKLADPRKPKMAKLPVILGNLDTIDNVSSDPVNY 599

```

```

Db 477 LLDGFSPLYKODSKSLEDILKLSKKREK-TKLQIFGQINITECVPLVLSNC 535
Qy 600 VNSSYPTQOF-TCSTPTTPEVEBFVPCIRPHQTPYITMNLVYKYLKYDSOKSF 658
Db 536 TSSVPLPFEKNCN--ITVEBEFVPEMTKYCPFTIYKHLVYPLQLYDSOKFE 593
Qy 659 AKARNIACIEFKDSDEDSQPLKCIYGRBGCVFTRSAFAVALHHONPEFYDEIKEL 718
Db 594 AKARNIACVEFEDSDSDASAKCIYGRBAGSVFTNNAVASHNQPFEYDEIKEL 653
Qy 719 PLOHEKHLTLTFHVSCDSSSKSTKRDVETOVGSWPLKXGCVTSEQH PVS 778
Db 654 PHLHOKHLTLTFHVSCDSSSKSTKRDVETOVGSWPLKXGCVTSEQH PVS 713
Qy 779 ANLPSCYLVQELGMRHNGPEIKWVDGKPKILKSTHLSVTVYTOODOLHNPQVCOCT 838
Db 714 ANLPGLYLNLDSESRQCNVDIKWVDGAPPLKFKSHLESTIYTODLHVHFFHQCQL 773
Qy 839 ESGAOLAGNELVYKLSLHAMEGHVIAFPTIINOLFVLTFRATOE-EVAVVETVYIH 897
Db 774 QSGSKVPEBELIKYLCCLHAMEIQWIOPLVYLMOLFVLTMTHEDDVPINCMTVLH 833
Qy 898 VVAQCHEBLSHLSYVYKAYAEPYVASEYKTVHEELTKSWTILKPSADFLTSNKL 957
Db 834 IYKSCHEEGDLYSFIKYSFPEKSPAPQOLIHETLATWIALKQSDAPLSINKL 893
Qy 958 KYSWPEFDVLIKSMAGHLNENSVKILRNORPPASVHHAVENTYNNLMHITQKFDNBE 1017
Db 894 KYSWPEFEIYANSMATVYLBENKIKLPQGRPEPEYHVAHSLILAIIPHVITRYAEID 953
Qy 1018 ASKNANSLAVFIKRCFTFMDRGVFKQINNYISCPADPKTLFEYKEEFLRVNCNHH 1077
Db 954 ESRNVVYSLASFILKRLTMDRGFINLINDYISGSPDPKTLAYKEFEIQTICNHH 1013
Qy 1078 YPLNLPMPGKRIQRYODLQDLSLTDEPCNHFVGLLREVGTALQEFREVLIAI 1137
Db 1014 YPLNLPMAFAKPKLQRYVQDSNLEYSLSDEYCHHVLGLLRETSIALQDNEYIRYTAI 1073
Qy 1138 SVLKLLIHSPPDRYASRSHORITVLPGLLIENVORINVDVSPFPNAGMYK 1197
Db 1074 SVLKLLIHAFAFTRYQHNQOAKIQLVLPVGLLENIQRLAGD----- 1120
Qy 1198 DESLALPAVPLVTPQKGSITLDSLHKDL-----GASIGIAPYTTSTPINSV 1247
Db 1121 -----TLXSCAMPASASRDEPCGFTSPANRGSISTOKRAYGFOGHCI 1167
Qy 1248 RNADSGSLISTDSGSLPERNSEKSNLDRHQOSSTLNSVVRCDLQSEIKSLMCF 1307
Db 1168 KREDSRGSIL-PEGATGFPDQGTGENT---RQST-RSSVSQVRLDQYEIRSLMCF 1221
Qy 1308 LVILKMSDALLFTYNNKASTSELMFPITSEVCLHQFOYMGKRYA----- 1354
Db 1222 LVIVKMSIDTLLTYNNKVSPOELINILILEVCLHFRMGKRNARVHADWLSKHGFI 1281
Qy 1355 -----RTGMHARLOOLGSLDNLSTFNHSGSDADVLHQSLEANIATEVCL 1402
Db 1282 DRKSOTMPALRNSSGMQARLOHLSLEBSFTLNHSTTTEADIFHQALLEGTAREVL 1341
Qy 1403 TALDLSLFTLAFKNQLADHGHNPLMKVFDVYICFLOKHQSEVALKNVFTALRSIYK 1462
Db 1342 TVLDITSFTQCFKTOILNNDGHNPLMKVFDIHLFLKNGQSEVALKHVFASTRFISK 1401
Qy 1463 PESTFEGRADMAALCYELIKCCNSKLSIRTEAQLIYFLMRNFPDTGKKSFPRTL 1522
Db 1402 PSBAFPGKRVNMCACVEVLKCTSKISSTRNEAALLYLLMRNFEYTKRTFTRLH 1461
Qy 1523 QVIVISQILADVIGITRFOOSLSIINNCAISDLIKHTSPSSVOKLTRIRFVLA 1582
Db 1462 QIIVASQILADVIGITRFOOSLSIINNCAISDLIKHTSPSSVOKLTRIRFVLA 1521
Qy 1583 TQMKHEHDEPMLVDLQYSLAKSYASTPELRKTWLDMARIHVKNGLSEAMCYVHT 1642

```

```

Db 1522 TQMKHEHDEPMLVDLQYSLAKSYASTPELRKTWLDMARIHVKNGLSEAMCYVHT 1581
Qy 1643 ALVAEYLRKEAVQNEPPLPHSHSACLRSGVGFROGCTAFRYITPNIDBEASMEVD 1702
Db 1582 ALVAEFLRKK-----LFGNCAFKKITINIDBEAMKEDA 1618
Qy 1703 GMQDVHFNEDVLMELLEOCADGLMAERYELIADLYKLIPIYERKDFERLAILYDTLH 1762
Db 1619 GMDVHYSEBEVLLEBQCVQGLMAKERYEITSEISKLIPIYERKREFEKLQYRTLH 1678
Qy 1763 RAYSVTEVMSGRLLGTFRVAFFGQAOYQFIDSETDVBGFPFEDDGKRYIKERKL 1822
Db 1679 GAYTILLEVMMHKRLGLTFRVAFYQGS-----FPEEDDGKRYIKERKL 1724
Qy 1823 TPLSEISORLKLKYSDFGSENVKMIQDSGVNPKDLSKAYIYOVTVIIPFDEKELOE 1882
Db 1725 TGLSEISRLVLYKEKGTENVKLIQSDKYNKELPKIAHIOVTVIYKPFDDKELTE 1784
Qy 1883 RTEFERSHNIRRFEMPEFTOTGRQGVBEQCKRRTILTAHCFPVYKRIIPWYQHH 1942
Db 1785 RTEFERHNNISRFFEAAPYTLGSKQOCIEQCKRRTILTSNSFPYVKRIIPINCEQ 1844
Qy 1943 TDLANIEVALDMSKKAVALROLCSSAAYDMIKOLKIQGSVSVVNAAGPLAYAFDLD 2002
Db 1845 INLKPIDATDEIKDKTAELOKCSSTDVDMIOLOLKQGWVSQVNAAGPLAYAFDLD 1904
Qy 2003 TMTKRYDPNKVYKLEVFROFPEACQALAVNERLIKEDOLEYOEBEMKANYREMAKELSE 2062
Db 1905 SQASKYPPKYSSELKOMRKRFIQACSTALBELERLIKEDQVEYHEGLSNFRDMKELSD 1964
Qy 2063 IMHEQICPLEEKTSVLP---NSLHIFNAISGT 2091
Db 1965 ITHQI--LQEDTHMSPMWSNTLHVFCALISGT 1994

```

```

RESULT 15
US-09-978-244A-29
Sequence 29, Application US/09978244A
Publication No. US2003010392A1
GENERAL INFORMATION:
APPLICANT: Lu, Peter S
APPLICANT: Garman, Jonathan D.
APPLICANT: Candia II, Albert F.
APPLICANT: Arbor Vita Corporation
TITLE OF INVENTION: CLASP MEMBRANE PROTEINS
FILE REFERENCE: 020554-000161US
CURRENT APPLICATION NUMBER: US/09/978,244A
PRIOR FILING DATE: 2001-10-15
PRIOR APPLICATION NUMBER: US 60/310,028
PRIOR FILING DATE: 2001-08-03
PRIOR APPLICATION NUMBER: US 09/737,246
PRIOR FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 09/736,969
PRIOR FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 09/736,960
PRIOR FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 09/736,968
PRIOR FILING DATE: 2000-12-13
PRIOR APPLICATION NUMBER: US 60/240,545
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,508
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,503
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,539
PRIOR FILING DATE: 2000-10-13
PRIOR APPLICATION NUMBER: US 60/240,543
PRIOR FILING DATE: 2000-10-13
Remaining Prior Application data removed - See file Wrapper or PALM.
NUMBER OF SEQ ID NOS: 106
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 29
LENGTH: 2008

```

TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: CLASP-4
US-09-978-244A-29

Query Match 55.8%; Score 6105.5; DB 9; Length 2008;
Best Local Similarity 57.8%; Pred. No. 0;
Matches 1197; Conservative 327; Mismatches 415; Indels 133; Gaps 24;

75 LPPYDFOALIRGCRITCTYPAKAEBAQSLPTTECITKYNSSMHLVNTKYEGSGE 134
1 MPPMEDISISVIGRORRTVOSTPBAERQSLPYKECITKYSDMHVNTKYBFSGD 60
135 PQLPKNVVKLKLPHVYEVDEVDKDDAASLGSGKGTGKGLYKGNMSAISYVM 194
61 PMLPEKSLRPEKIPNHVEIDEDCEKDDSSSLCGKGGVTKQGLHAAVNSTITVTM 120
195 RSPKGRFFHLIQLDGGSYKFEFLKDLQ--KEPKGSIFLGLYGVSPFNKVRPAFELM 252
121 KVFERRFYLTQLPDGSYILNSYKDEKNSKSGCIYLDACIDV--VQCFMRHAFELM 179
253 QKSSVTLADSEVEHEWITLTKLQLNFAAMEKRG--DSHEDD---QSKLESG 308
180 LDKYSHYLAETEOEMEWLITLKKLIQINTDSLVEKKIYETADDBETSSQGAENIM 239
309 SGLD--SYLPFLAKSAREAE--IKL--KSBSRVKLFYLDPAOKLDFSAEPYKSPFEKFG 364
240 ALEBSMPELMKYGHETQNLKLSGDRONLFSFDSEYORLDSGIEPDIKPEEKN 299
365 KRLVKNDSLSPNLQCCVANESEPTTNEPFFVTLSPDIKYNKISADFYDLNHFVS 424
300 KQFLVNCBTLTFLNIQLGIDNAKGPRTVEPFINALFDVNKNCKISADFYDLNPSV 359
425 RQMIATSPALNM-----GSGPETOALAGILHEAMQYKPGISVYTPHEDIFLVARI 479
360 RMLMGSSVQLASDGS PKSSPE--SYHGIA--ESGLRYIOGIFSVTPHBEIFLVARI 416
480 EKYLGOSITHCAEPYKSSDSKVAOKVJKNAKOACORLGOYRMPAMARLTFKASGN 539
417 EYVLQGNITHCAEPYKNSDPVKYAKOVRTAKOVCSRLGOYRMPAMARLPFKDTQGS 476
540 LDKNARFSAIYRQDSNKLSDNDMLKLLADFRKREKXAKLPVILGNLITIDNVSDFPNY 599
477 LLDGRFSPLYKQDSSKLSBIDLKLSBYKKPEK--TKLQIIPGQNTIIVECPYDLSNC 535
600 VNSVITPTKQFE--TCSKPTITFEVEEFVPCIPKHTOPYITINMLVYPRKYLKQDSQSF 658
536 ITSSVYPLRPFKNCON--ITVEVEEFVPEMTKYCYRPTIKYKHLVYVPLQKYSQKTF 593
659 AKARNIACIEFKPDSDEBDSOPLKCIYGRPGGVPTRSAFAAVLHHNHONPEFYDEIKEL 718
594 AAKARNIACVEFRDSDSDASALKCIYGRPGASVFTTNAAVASHNQPPEFYDEIKEL 653
719 PTOLHEKHLTLTFPHVSCDSSKSGTKRDVVEITQVGSWLPFLKDGQVVTSEQHPIVS 778
654 PHLHQKHLLTFPHVSCBINTKGTTKQDVTETVPGFAWPLKDGKIIIFEQDLPS 713
779 AULPSCYGLQYOLGMRHAYGPEIKAVDGGKPLKISTHUVSTVYTYDDOGLHNPFOYQCT 838
714 AULPPLYMLNDAESRQCNVDIKWVDGAKPLKFKSHLESTIYTDLHVHFKFPHCQDL 773
839 BSGAOLAGNELVYKLSLHAMEGHVIAFLPTILNQLFRVLIRATOE--EVAVNVTVIHH 897
774 QGSGKEVPBELIKYLKCLHAMEIQVNIQPLPVILMQLFRVLINMTHEDDVPINCTWVLAH 833
898 VVAQCHESGLBSHLRSYVYKAYKABPYVASEYTVHEELTKSMTIILKPSADFLTSNKL 957
834 IYKCHESGLDYSRLFYKSPREKSAPOAQLIHELATMIMAILKQSAFLSINKL 893
958 KYSWPFVDVLISMAOHLIENSKVLLRNQRPASHTIAVEVYVNMMLPHITQKFRDNE 1017
894 KYSWPFELIAKSMATYLLBENKIKLPRQGRPEPTVHVHLSILLALIPHVTIRVAEIPD 953

1018 ASKNANSLAVETIKRCTFMDRGFVFKOINNYISCFAGDGPXTLEFYKFEFLVVCNHH 1077
954 ESRNNVYSLASTARCLTMDRGFTFNLINDYISGFSKDPKVLAVKPEPLQTCINHH 1013
1078 YIPLNLMPFGKRIQRYVQDQLQDLYSTDEFCNHFVGLLREVTALQDFEVRLLAI 1137
1014 YIPLNLPMAPAKPQLQFVQDSNLEYSLEDECKHFLVGLLRETSIALQDYETRYPAL 1073
1138 SVLKLLIKHSPDRDYARSNOARLATYLPFLGILINNOVINRVDSPFVNAGMTVK 1197
1074 SVLKLLIKHAFDTRYQHKNOQAKLAQLYLPFVGLLREINRLARD----- 1120
1198 DESIALPAMNPLVTPQKSTLDNSLHKDL-----GASIGIAPYTSTPNINSV 1247
1121 -----TLYSCAMPNNSRDEPFCGPTSPANRBSLSTDKPTAGSPNGHGI 1167
1248 RNADSRGSLISTDGSNLSPERNSEKNSLDKHOSSYTGNSVVRCDKLDQSEIKSLMCF 1307
1168 KREDRGSLLI--PEGATGFPDQNGTEGNT-----RQST--RSBSVSYNRLDQYEIRSLMXY 1221
1308 LYIILKMSDDDLFTYWNKASTSELMDFITISBVCLHOGYNGKRYIA----- 1354
1222 LYIVMISEDTLTYWNKVSQDELINILILEVCLFHRYGKRIAVHDAWLSKHFGI 1281
1355 -----RTGMHARLQOLGSLDNSLTFNHSYGHSDADVHQSLLEANIATRYCL 1402
1282 DRKSQTMALNRSSVMAQRIQHLSSLESSPFLNHSSTTERADIHQALLBENTATYVSL 1341
1403 TALDLSLFTLAFKNQLLADHGNPLMKVVDVYLCPFLQKIQSESTALANVPTALSILYK 1462
1342 TVLDITISFPTQCFKQQLNNDGNHPLMKVFDHILAFKNGQSEVSLGHVFPASLAFISK 1401
1463 PPSFTYBGRADMCALCTEILKCNKLSIRTEASQLLYPLMRNPNYTGKSVRTHL 1522
1402 PPSAFKRVNMCAPCYEVKCTCSKISSTRNEASALLYLMRNPFYTKRKTFLRTHL 1461
1523 QYIISQULIADVIGIGTRFOQSLIINNANSRDLIKHTSPSSDVLDTKRIRTYMA 1582
1462 QITIVSQULIADVALSGRFQESLFTINNANSRPMKATAPRAVEVDLTKRITVMA 1521
1583 TAQMEKENDEPMLVDLQYSLAKSYASTPELRKTLWDSMARIHVNGSLSEAMCYVHT 1642
1522 TAQMEHEKDEPMLDLYSLAKSYASTPELRKTLWDSMAKIHVNGSPSEAMCYVHA 1581
1643 ALVAEYLTRKAVQMEPPLRHSNACLRBSRGVFRGCTAPRYITRNIDSEAMMEDV 1702
1582 ALVAEFLRRKK-----LFPNGCSAFKKITENIDSEAMKEDA 1618
1703 GMQDVHFNEDVLMELLBOCADGLWKAERYELIADLYKLIIPYKGRDPBRALHLYDTLH 1762
1619 GMQDVHSEBVLBLECCVDGLWKAERYELISELSKLIYIYKGRFEBKLTQVYRILH 1678
1763 RAYSRYTEVMHSGRRLGTFRVVAFFGQAQYQFTSETDVEGPFSEDEGKEYIYKPEPL 1822
1679 GAYTILFVMTYKRRKLGTFPRVAVYQGS-----PFEBSDGKEYIYKPEPL 1724
1823 TPLESISQRLKLSYDKSGSENVKMTIQSGKVPKDLDSKAYIYOYTHVIPFDEKELOE 1882
1725 TGLSEISRLVLYLGEKGTENVKIIDS DYNAKELPKTAHIOYTVKPYFDDKELE 1784
1883 RKTPEFSRSHNIRBPFMEPFTOTGRQGVGEQCKRRTILAIHCFPVYKRIIPMYOH 1942
1785 RKTPEFSRSHNIRBPFMEPFTOTGRQGVGEQCKRRTILITNSFPYVYKRIIPINCEQ 1844
1943 TDLNIEVALIDEMSKVAVELQOLCSSAEVMIKLOLQKQSVSVVGNAPLAVARAFIDD 2002
1845 INLKPIDATDEIKDKTAEQLKCSSTVDVMIQOLKQGVNVSQVGNAPLAVARAFIIND 1904
2003 TINTKAYPNNKYLKLVKAVROVEACGALANVERLIKEDOLEYQHEMAQNTRENAKELSE 2062
1905 SQASKYPPKQVSELKDMFRKPIQACSIALBELNERLIKEDOVEYHGLKSNFRDVKELSD 1964

Mon Jul 14 19:53:42 2003

us-09-815-379-8.rapb

Page 20

```

QY      2063  IMHEQICEPEKTSVLP--NSLHFNIAISGT  2091
          |::|||::|::|::|::|::|::|::|::|
DB      1965  ITHQEI--LQEDTMHSPWMSNTLHVFCAISGT  1994

```

Search completed: July 14, 2003, 18:27:11
Job time : 59.3964 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:10:22 ; Search time 46.6135 Seconds
(without alignments)
4345.419 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936
Sequence: 1 MSQPLPLPASAETKRRKTRAL.....ISGTPSTWVHGHTSSSSVV 2107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2888	26.4	738	2 I60486	gene tlg protein -
2	1977.5	18.1	2374	2 T31052	hypothetical prote
3	1664.5	15.2	2018	2 T34274	hypothetical prote
4	540	4.9	1105	2 A71430	hypothetical prote
5	429.5	3.9	1759	2 T18868	myoblast city prot
6	382	3.5	1970	2 T03284	myoblast city prot
7	279.5	2.6	2550	2 B53435	vesicular transport
8	232.5	2.1	1932	2 S53409	probable membrane
9	223.5	2.1	2819	2 A90551	conserved hypothet
10	224.5	2.1	831	2 T08611	hypothetical prote
11	222	2.0	2166	2 G70163	hypothetical prote
12	224	2.0	2401	2 T28676	rhodopsin protein -
13	213.5	2.0	3685	1 A27605	dystrophin, muscle
14	212.5	1.9	2954	2 T14156	kinesin-related pr
15	204.5	1.9	843	2 T01438	hypothetical prote
16	203	1.9	2077	1 W2B224	240K tegument prot
17	198.5	1.8	1620	2 S61535	nucleotide-binding
18	197	1.8	2028	2 T08025	DNA-directed RNA p
19	195	1.8	2078	2 T09326	tegument protein -
20	195	1.8	3599	2 P90608	ABC transporter pe
21	194	1.8	3119	2 T18414	protein g377 - mal
22	193	1.8	1325	2 T42722	male-enhanced anti
23	192	1.8	3225	2 I52300	giantin - human
24	191	1.7	3259	1 A56539	giantin - human
25	189.5	1.7	2748	2 S57976	nuclear migration
26	189	1.7	2269	2 T28677	rhodopsin protein -
27	189	1.7	2297	2 AB2494	hypothetical prote
28	188.5	1.7	1939	2 T18372	repeat organellar
29	187.5	1.7	1778	2 JT0382	apolipoprotein B -

30	187.5	1.7	2629	2 I46569	apolipoprotein B -
31	187.5	1.7	2829	2 A42771	reticulocyte-bind
32	186	1.7	1818	1 S73852	hypothetical prote
33	186	1.7	2663	1 S28261	centromere protein
34	185	1.7	1557	2 T18412	lipid-binding prot
35	185	1.7	4464	2 D87755	protein T1812.4 l
36	184	1.7	840	2 T01357	hypothetical prote
37	184	1.7	1289	2 S67200	hypothetical prote
38	184	1.7	4385	2 T29042	hypothetical prote
39	183.5	1.7	1957	2 T38077	hypothetical coile
40	181	1.7	1837	2 T41023	probable nuclear p
41	181	1.7	2014	2 S46622	probable membrane
42	181	1.7	2649	2 A40937	bulbos pemphigoid
43	179.5	1.6	1963	1 MWKM	myosin heavy chain
44	179.5	1.6	1997	2 F71607	DNA helicase II Br
45	178.5	1.6	1211	2 S68251	phospholipase C, 1

ALIGNMENTS

RESULT 1
160486
gene tlg protein - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
C:Accession: I60486
R:Pinase, L.; Porcellini, A.; Avvedimento, V.E.; D'Esposito, F.; Feliciello, A.; Montice
Mol. Biol. 13, 75-83, 1994
A:Title: A novel thyroid transcript negatively regulated by tsh.
A:Reference number: I60486
A:Accession: I60486
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-738 <RES>
A:Cross-References: EMBL:X68101, NID:G550419, PIDN:CAA48220.1, PID:G550420
C:Genetics:
A:Gene: tlg

Query Match	26.4%	Score 2888	DB 2	Length 738
Best Local Similarity	87.1%	Pred. No. 9.8e-159		
Matches	573	Conservative	23	Mismatches 36; Indels 26; Gaps 7
QY	1424	GNPLMKVFDVYLCLFQKHQSEPLAKVFPALSLIKFPSTFEGRADMCALCYEIL	1483	
DB	5	GHSLMKVFDVYLCLFQKHQSEPLAKVFPALSLIKFPSTFEGRADMCALCYEIL	64	
QY	1484	KCCNSKLSIRTEASQLLYPLMRNFDYTGKSPVRYHLOVYISVSQIADVVGIGTRF	1543	
DB	65	KCCNSKLSIRTEASQLLYPLMRNFDYTGKSPVRYHLOVYISVSQIADVVGIGTRF	124	
QY	1544	QOOSLIINNCANSDRLIHGTSFSSVYKDLTKIRIVLMATQMKHEHNDPEMLDLOYSL	1603	
DB	125	QOOSLIINNCANSDRLIHGTSFSSVYKDLTKIRIVLMATQMKHEHNDPEMLDLOYSL	184	
QY	1604	AKSVASPELRKTMDSMARIVHKNKGDLSBAKCVHATVAAYLTTRKE--AVQWBP	1660	
DB	185	AKSVASPELRKTMDSMARIVHKNKGDLSBAKCVHATVAAYLTTRKEADLAQREBP	244	
QY	1661	LLPHSHACLRSSRGVROGCTAFRVITPINIDEASNMEDVGMQDVAFNEDVLMELLEQ	1720	
DB	245	VFPYSHTSQCRKSRGCMRGCTAFRVITPINIDEASNMEDVGMQDVAFNEDVLMELLEQ	304	
QY	1721	CADGIMKAEKRYE--LIADYKLIIPYKRDPEFLAHLYTTLHRAYSKVTEFWHSGRL	1778	
DB	305	CADGIMKAEKRYE--LIADYKLIIPYKRDPEFLAHLYTTLHRAYSKVTEFWHSGRL	359	
QY	1779	LGTYPRV--AFPGQAQYQFTDSEVDGEPFEDDGKXYIKKPLQPLSEISQRLKXY	1836	
DB	360	AGSMPLPGSLPGQ-----GPFEDDGKXYIKKPLQPLSEISQRLKXY	405	
QY	1837	SDKFSSENVKMIQDSGVNPKDLSKAYIVQTVHVPFDEKELQERKTEFERSHNRFP	1896	

Db 406 SDKFSSENVKMIQDSKVNPKOLDSKFAVIQVTHVTFPEDEKELQSRKTEFERCHNIRRP 465
Qy 1897 MEMEPTQTGKSGQGVGEQCKRTTILTAIHCFPYVKRIIPVMYQHHTDNLPIEVALIDENS 1956
Db 466 MEMEPTQTGKSGQGVGEQCKRTTILTAIHCFPYVKRIIPVMYQHHTDNLPIEVALIDENS 525
Qy 1957 KKVAEIROLCSAEVDMIKLQKLGSSVSQVNAAGLAAARAFLDJDTNTRKYPDNKVLL 2016
Db 526 KKVAEIROLCSAEVDMIKLQKLGSSVSQVNAAGLAAARAFLDJDTNTRKYPDNKVLL 585
Qy 2017 KEVFRQFVACCGALAVNERLLIKEDOLEVOEEMKANYREMAKLSIMEQICPLEEK 2074
Db 586 KEVFRQFVACCGALAVNERLLIKEDOLEVOEEMKANYREIRKELSDIIVPRICPGBDK 643

RESULT 2

T21052
hypothetical protein F22G12.5 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_rev1510n 15-Oct-1999 #ext_change 29-Oct-1999
C:Accession: T21052; T21281
R:Kerhaw, J.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19365
A:Accession: T21052
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2374 <M11>
A:Cross-references: EMBL:Z81066; PIDN:CAM02974.1; GSPDB:GN00019; CESP:F22G12.5
A:Experimental source: clone F17B5
R:Lenhard, N.
submitted to the EMBL Data Library, March 1997
A:Reference number: Z19400
A:Accession: T21281
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2374 <M12>
A:Cross-references: EMBL:Z92831; PIDN:CAM07369.1; GSPDB:GN00019; CESP:F22G12.5
A:Experimental source: clone F22G12
C:Genetics:
A:Gene: CESP:F22G12.5
A:Map position: 1
A:Intron: 47/2; 75/3; 114/3; 172/2; 230/3; 322/1; 376/2; 439/1; 653/2; 731/3; 758/1; 80
/2; 2139/1; 2258/3; 2298/2; 2344/3

Query Match 18.1%; Score 1977.5; DB 2; Length 2374;
Best Local Similarity 25.6%; Pred. No. 2.4e-105;
Matches 628; Conservative 416; Mismatches 918; Indels 489; Gaps 73;

Qy 9 ASAEKTRKTRALSKPTAAELROSSEVVGSVLAKPLIEPLDYENVIQKQTQILND 68
Db 41 ATSQARKFTAGVGRPELAEARDAVAALVAQNPKEPISSEVYVNEKFINERSIQLEND 100
Qy 69 CUREMILFFYDFQTAIILROGRYICSTPAKAEBAQSLPTECIKTYSNMHLVNYK 128
Db 101 POREIYLFPRDIEESQOTCEVAPVYKNDIKAKMLTREARFYTTAHSRISVNY 160
Qy 129 EDYSGEF-RQLPKVYKVLKLPVHYVEVDVEDKDASLSQSGKGTGKGLYKGNM 187
Db 161 SKFSGDYSAANTSSAIENETLSLVEESDALE-DEKAAFGASSNCAG---GVVEGGLN 216
Qy 188 SAIS-----VTMSFKRFRPHLQLDGSYKFEFLDQKEPKGSIFLGLYGVSPR 239
Db 217 VLKSDPPLTFDNLKSSSARRYCILRHGGECTFIRKTAQEAR-MLPLKVAQAQK 274
Qy 240 NKKVRFAELKQ--DKSYLLADSEVMEEMITL----- 275
Db 275 STRKGTVEIRGSDADKAAIILLESBETQMLEMVLVAHTAALANKEDAISICSELDYK 334
Qy 276 ---NKILQINFEAAQMEKNGSDHEDEQSKLEGSGGLDYLPELAKSAREAEITLKS- 331
Db 335 GIGSKLIDVNAATSSSTVGGGRSKGADTESIGSEDSSNCW-----NSANFEPKKLKNC 389

Qy 332 ---ESRVKL-FYLDPAQKLDPS---AEPEVKSFEKFGK-RIL----- 368
Db 390 TFFFRKRLDIFYSKTRKSPFSEBNLKFEKIQYIKFKGKTLIPFTRTSEMQWRG 449
Qy 369 -----V 369
Db 450 RNSAARALQPIVDRRMNFSLYRLQPLQSPHESSTILPGNSKRGADAPIFSSMSPV 509
Qy 370 KCONDSPNLOCCVAB-----NBEGPTVNEPFFVTLSPD 404
Db 510 KMRKTSGLKRGKGAOSSLSISTGTFFLSLDPHGLHLPLNLSGLVLOQIEFFVFAFPFD 569
Qy 405 IKYNRKISADPEVNDLHNSVQMTATTSPALMNGSGPETOSALRGILHEAAQYKQIGF 464
Db 570 ALFARRACBEFQILVN---GDELHLPVF--DSCKYNIGIRRLVDRASRALQLP 623
Qy 465 SVTCHPDIPLVAREKVLQGSITTCAPPYKSSSSVQAQVUKNAQACORLQYRMP 524
Db 624 PTTVNRDLMTVCYCRDRL--SADTAELYNKSSDPRAVAKLOQTVSMSLTRLGHRQR 681
Qy 525 FAMAARTPEKASGNLDKNA---RPSAIYRQDSNKLSDNMLKTLADPRKPEKNAKLPYI 581
Db 662 FAWTKAPLPPELRSVTHQRSARRRLCFLHQLDQOKYQOML--FNCTDRQSKMILP 739
Qy 582 LGNLDITIDNVS--DFPNYVNSYIPTKQF--ETCSKTPITFEVEEFVPCI PKHTQPYT 637
Db 740 NATVSVSVTRASVIVLYNVRVPSLYPNPKBEDPSTMPVFGQSGF---DQSGQPH 796
Qy 638 IYTNHLYVYPKTKDKDSQKSPAKANLACLEFQDSDEDEQPLKICYR--PGFPVYTR 695
Db 797 SLTNLLYIYPLALKDSQKAFSKANISCTYRFRNG--EEAIPKAMVDRMSAGP-YCI 853
Qy 636 SAFAAVLHHQNPFEYDIEKIELPTQHEKHNLLTFPHVACDSSKSTKRDVETQV 755
Db 854 SSTCAVQHQQNPVGEEMKQQLPLNLTSHLPFSHISVAGNS--NLKASESTETPI 911
Qy 756 GYSWLPLI--KGRVVTSEQH--IPVANSPLSGYGYQVGLMGR--HYGEPIKVVDSK 808
Db 912 GYSWLPLWKCDRLVMEDEQFALPVADIPPNYVRSKPTGKMGDETLTSEVRVWD--QK 970
Qy 809 PLKSTHLSVTVYQDHLNFPQYCKQTSAGNALGNELVTKYLSLHAEHGMIMFL 868
Db 971 PLFERRLVSSVFTFDKDLTFQACQRL--SAGGIIIGDAADSKLSDVPRDLVYL 1028
Qy 869 FTILNQLFRVLTRAQOEAVANVTFRVLIHVVAQCHEBGLJESHSYVKAAY---APY 924
Db 1029 PIVGLRLALIPQATEGLATATLRSISICDMCCQNKQPVVRRFVSHSDSARQF 1088
Qy 925 VASEYKTVHEBLTYSMTILKPSADFLTSN-----KLKYSWFPEFVLIKSMAOHLIENS 979
Db 1089 VSHL-TTIVSAICRHLPLMBE---LQSEPEQLAQVLYQLVLTDAVYKSMQAQMSQEQ 1143
Qy 980 KKKLIRNQRFPASVTHAVETVVMMLPHITQKFRDNPASAGNHSLSLVFIKCTFMDR 1039
Db 1144 LNKTSQRORFQOELLQGOVLEGVPOIVLKHMEKREBSCANLALAYFTRFAMSFDVR 1203
Qy 1040 GFVFKQINNYISCFAPGDKPTLFEYKFEFLVVCNHEHYIPLNLP----- 1085
Db 1204 GVFWRHIFYSRLDDTDYRALDYKTDLBILCLHEHNVANLPVLLNAAQIORLAYS 1263
Qy 1086 -----PFGKG-----RIQ 1093
Db 1264 GGVVDVTQMTNGTNGLSGGYCSLSRSRTVVVNNIRASGSGFLSRFENQIFNTPLETNETD 1323
Qy 1094 RYQDLQDYSLTDECRNHLPLVGLLRVGRALQEFRRVRLAISVLYKLLIKSFPDRY 1153
Db 1324 RYASGSGEMHLSPAFQNHFTVGLMOELVACIRBTDYRRKRPISLNLNLLAKSHFDKRY 1383
Qy 1154 ASRSHQAIATLYPLFGLILEINJORINVRVS-----PEPVNAGMTVKDESIAL 1203
Db 1384 GDMTQIRIMLYAPVLTALFDHLHERPSEBIDVDVATPTGYRSPTSGGRLP----- 1436
Qy 1204 PAVNPLVTPQKSGTLDNSLHKDLGALSGIASPYTTSTPININSVNNADRSGLISTD--- 1260

```

Db      1437 PSSSTTKTLKSKRTYVR--YTDVQPSRG--SPIRTSSP--VPSQTTPSRPPQMPPPPPA 1491
Qy      1261 SGNLSLPERNSEKSNLSDKHQSSSTLGNVRCQDQSEIKSLMCFLLTKMSDDALF 1320
Db      1492 SSKTTPPPPS-----SSTPLVEKLTEDBEIDPLICCVTLQMPKRIIA 1535
Qy      1321 TYW--NKASTSE-----LMDPF--TISEVCLHQP-----QYMGKRYAR 1355
Db      1536 ALMTBNDGANNAEMKRLRLIYDVPLFEMALQNGVIEYVGKIMLVLKRRKRGKHNAR 1595
Qy      1356 TGMHARLQOL-----GSLDNLST---FNHSGHSDAD-----VLHOSLIE 1393
Db      1596 RTAANSKTRSLFTLNLPSRISSGSTPNRISAEWMAAGAEEDDPNNGSTTNPIPFVTLQ 1655
Qy      1394 ANIATEVCITALDTLSLFLAPNQLAD--HGHNPF-----LMKKVVDVYLCFLQKQSQST 1447
Db      1656 VNLSEVALIYLDVAO---TPAHQLASSQRHRCQSDALFHSLLALHLRLADEHWSST 1711
Qy      1448 ALKNVFTALRSLLYKPPSTFYE--GRADMCAALCYEILKCNKSLASIRTEASQLYFLMR 1506
Db      1712 VRLHYTAGALFVNLFBRARLFEGGPLEPLYMILEKVLQMASRLPAVQAAAAALLQLILR 1771
Qy      1507 NNPD---YTGKKSFVTRLQVITISVQLADVNGIGTR-----PQQ 1545
Db      1772 NGYEVAOQYFAQVLAQSVSPSSSKINQQAATRKGVSSERLGRPGSQGVALARLLGFC 1830
Qy      1546 SLSTIINNCANSRLI-----KGTSPSDVDQDTRKIRTVLMTATQMKHEHNDREM 1595
Db      1831 --SVLSNSAPFERGLAAVEALVDTRKATSFDAVLDLRLQGVMTATVATLQDAANDPFR 1888
Qy      1596 LVDLOSLAKSYASTFELRKTMLDSMARIVKNGDLSAAMCYVHTALVAEYL--TRKE- 1653
Db      1889 LADLHQLADSVRGSAAALSAMFDTLAEIYEDRWAEASVCHASVALIAELERKEL 1948
Qy      1654 AVQMEPRLPHSHSACLRRSGVFRQGTAFRVITPNI DE--BASMEDVG--MDVHENE 1711
Db      1949 EVDWR-----VPMINRIJAETEOSGCGAGSVQPAFGTT 1983
Qy      1712 DVLMLLEOCADGLMKEARYELLADIYKLIPIYERKRPERLAHLYDLTHAAYSVTEV 1771
Db      1984 DNLGAKIDTAAALMLAEKFEAVGPLYRLIVPLEKNMIFTSLSVSYAELQOTYSRAAEV 2043
Qy      1772 MHSGRLLGTYYRVAVFQQAQYQFTDSETVGFEDEGKEYIYKEPKLTPLSEISQR 1831
Db      2044 RSSGKRHLGAYFRVRNG-----ERNHGSHTNDWIYREGSLISLAFALE 2089
Qy      1832 LKLKYSDFKGSBNVKMIQDSKVNPRDLSKYAYIQVTAVIPFEDEKE--LQERKTEPER 1889
Db      2090 IREKQORQVGHDRVQ--IEANEOILSKIDPTVAVYQITFVEPSIPAAAGIADQHRNDFLV 2148
Qy      1890. SHNIRFEMEMFTQYK--RQGVREOCERTIL-----TAHCFPPYKRI 1935
Db      2149 HTNLSEFSECATTEERKVSKEPAHEQCLKRTYLRVSPSVSEDSRAATGFPATRRLL 2208
Qy      1936 PYMVOHTDNLPIEVAIDEMSKVAELROLCSA---EVDMIKLOLTKQGSVSVQVNG 1991
Db      2209 PVISVHPBQSPLEFRACQCLTKABOIKRTIINAASGRQLDVKGLOLLQSAVLPTVNG 2268
Qy      1992 PLAYAVAFLDITNTKYRYPDNKYKLKEVFRQVEACGALAVERLIKEDOLEYOEMKA 2051
Db      2269 PLAYAEVFTKEQREREYGDGLVKLRSEFRNLNNSCOLAIEANASAIQSDQTYHEVLVS 2328
Qy      2052 NYREMAKELSEIMHOICPL-----EKTISVL--NSLHIFNALSQ 2090
Db      2329 SF-----DANHERLQTFPGASLRSLEPDSSSQFPPRSAMHILDMWG 2371

```

RESULT 3
T34274
hypothetical protein F46H5.4 - Caenorhabditis elegans
C1:Species: Caenorhabditis elegans
C1:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000

```

C1:Accession: T34274
R1:Nhan, M.
submitted to the EMBL Data Library, November 1995
A1:Description: The sequence of C. elegans coemid F46H5.
A1:Reference number: 221498
A1:Accession: T34274
A1:Status: preliminary; translated from GB/EMBL/DBJ
A1:Molecule type: DNA
A1:Residues: 1-2018 <NHA>
A1:Cross-references: EMBL:U41543; PIDN:AA837023.1; GSPDB:GN00028; CESP:F46H5.4
A1:Experimental source: strain Bristol N2; clone F46H5
C1:Gene: CESP:F46H5.4
A1:Map position: X
A1:Introns: 16/2; 52/3; 87/2; 116/2; 138/2; 203/1; 265/3; 317/2; 337/3; 378/1; 428/1; 482/3; 1491/3; 1560/2; 1632/2; 1753/3; 1830/2; 1866/2; 1927/3

Query Match      15.2%; Score 1664.5; DB 2; Length 2018;
Best Local Similarity 23.1%; Pred. No. 2.2e-87;
Matches 546; Conservative 406; Mismatches 746; Indels 665; Gaps 73;

Qy      8 PASAETRTALSKRTPAELROSV-----SEVVG-----SVLAKPEL 48
Db      3 PADGK-RAPVKKRSIV-TASVKKRVISGLPHIRLSEGDGLDAVSMMEKIQTEP-- 58
Qy      49 IEPLEENVIVQKKTQIINDCLREML-----LFPY--DDFOATILRQO--RYICST 96
Db      59 VPPVVEALLDQRT-----CLVYSSPKSKPLFEVSDVAVTTIRQDGLTDHNTT 113
Qy      97 VPAAKEEBAQSLFTECIKTYNSDMHLYNKYEDYSGEFPQLPNKVAKLKVHVEVD 156
Db      114 IP-NIENH-----VADIGFCYCDNPSLVNRYK----- 139
Qy      157 EKVDDDEDAASISGQKGTGKMGLYKGNMNASISVTMRSPRRRPFHLIQLDGSYKPEF 216
Db      140 -----AQGTED-----IKRMLEKLTALRSLLPQLFH-----NG 170
Qy      217 LKDLQEKPKGSIFLFLGVSEFRNNKVRFAFE-----LKMQ--LKSSYLLAADSVEY 268
Db      171 LRIMREP-STISID---GGSFTDIDPTKCKATEQYVSMLKSAIDKLTLYSMLSKRV 226
Qy      269 EEWITLNLKIQLNFEAAMQEKRNDSHEDEQSLSESGSGLDLSYELPILAKSAREATIK 328
Db      227 KIFMNCIN-----EEEKSWKMERKQ-----VPLRP----- 251
Qy      329 LKSESRVLFYIDPDAQLDPSAEPYKSPBEKKGKILVKCNDLSRNLQCCVABNEBG 388
Db      252 -EQEETPKLFLV-----KVERKAADP----- 270
Qy      389 PTTNVEPPTVLSLFDIKYKRNKISADPFVNDLHNSFVRQMIATTSPALNNGSGPTQSALR 448
Db      271 ---FPEPLPASMATIDIKORQKVTSMTFNADHKLDMLSHQHPFNIN----- 318
Qy      449 GILHEAMQYPRKQGISYTCRHPDIFLVAIRIEKVL--QGSITHCAPYKNSDSSKYAOKV 507
Db      319 -----MQVLFNVTGKLEDMFLVVKIEKVLQGDVDFENSPPYGTQDENNM--EKL 366
Qy      508 LKNAQAQCORLQGYHMPRW----- 527
Db      367 EBAERKNCORLGAAYPLGFOVIDLQRIYKANVSTGASSPFRRTDPMMSQCTTASGAVL 426
Qy      528 -----AARTLFPKDAEGNLDK----- 542
Db      427 TTAGSQSDQDQGSITSADRTIASMSGSLRRSGSALIVFSRVVTPLTLLKRFARVSNLP 486
Qy      543 -----NAFSAIYRQDSNKLNDMDLKLADFRKTEKRAKLPVILGNLDI 587
Db      487 TSGQVPEINEMWPSCNLKFSSPIRQEGDQSDDEDIYRICEKRRKNGVKHKKMFLELEL 546
Qy      588 TI--DNVSSDFPNVYNSSYIPKQPEKTSKPIPEVEVEFPFCIKKHQPTI-ITNNILY 644
Db      547 TLGANSKKEVQSHGSLN-----TLNSERVIHAEW-----IPVYASLNSKYQNVIF 594

```

[illegible]

```

Db      1540 EMTIDMLQYVEGSSNNPDLRITWLLMABRHEQRLCPBAHSYLOASLVEYIYIQO 1599
Qy      1554 A-VQMEPPLLPHSHSACLRSGRGVFRQCTAFRVITPNIDEA-----SMMEVDGM 1704
      1560 QNLSFE-----SKGATFEIITPNAIKESKTIPNSLKNADSENHI 1639
Qy      1705 QDVAFNEDVLMELLQCADGMKAEKREILADIYKLIPIYEXKRDEPRLAHLDTLHRA 1764
      1640 QSYHTEAGVAKILEKAPALLEKQVLELLPEPSKIILKTYHATKSSRASHHKKRGIA 1699
Qy      1765 YSKYTE-----VMSGRRLLGTFRVAFPGQAAQVQPTDSETDVEGFED 1809
      1700 ADQIKETGEYYENSODAMISPLPIDIDRCGSGFFRVAFYK-----LFGA 1744
Qy      1810 EDGKEYIYKEBKLTPLSEISQRLIKLYSDKFGSENVMMIDGSKNPKDLSKYAYIQVT 1869
      1745 LNNAEFYKESAFSKLNEISNRLETFYTMNGEGENVVLLKDSKPVQLEKLNPEKAYIQIT 1804
Qy      1870 HVIFPFQBEKLEQEKTEFEFSHNIRRMFEMPFIQYQSGKRGQVEQCKRTITLAINCFP 1929
      1805 FVDVYLLSDNEMEMETTYFTFRNNNNNREYFPAPYMEBGRAGGELAAQYKKKTTILVENSFP 1864
Qy      1930 YVKRIFPMYQOHNDLNPRIEVAIDEMSKYAEILQOLCSAEBDMIKYQLKLOGSVSVOVN 1989
      1865 YIKRRLQVNNRSVDQESPIEVALIEDIKKTRBELSAAOHQNPKN--LSMFIQSSIGTVN 1922
Qy      1990 AGPLAYARAF-----LDDTYTKRYPDNKVKLLEKVEFRQFYEACGQALANVERLIKEDOLE 2044
      1923 QGPLEIANVFANAMLMDRG--RPVDELQNKRLISFHLQCKAMEAIETLCROLTIGEDQKE 1980
Qy      2045 YQEMKANRYEMAKELSEIMHEQ 2067
      1981 YQRVNEENFESFVTHLKPMLSRQ 2003
Db

```

RESULT 4
A171430
hypothetical protein - Arabidopsis thaliana
C|Species: Arabidopsis thaliana (mouse-ear cress)
A|Variety: Columbia
C|Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_update 05-Dec-1998
C|Accession: A171430
R|Beyan, M.; Bancroft, I.; Bent, B.; Love, K.; Goodman, H.; Dean, C.; Berkamp, R.; Dijk
P.; Weller, H.; Wedder, E.; Wambutt, R.; Weitzensager, T.; Pohl, T.M.; Terry, N.; Gietzen
avansch, T.; Hempel, S.; Kotter, P.; Entlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.
Nature 391, 485-488, 1998
A|Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech
erholt, A.; Moore, T.; Jones, J.D.G.; Blevins, T.; Palmer, K.; Benes, V.; Reuchan, S.; Ansa
C.; Chatawal, N.
A|Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis thal
A|Reference number: A171400; MUID:98121113; PMID:9461215
A|Accession: A171430
A|Status: preliminary; nucleic acid sequence not shown; translation not shown
A|Molecule type: DNA
A|Residues: 1-1105 <REV>
A|Cross-References: GB:257340; NID:g2244950; PID:e327501; PID:g2244990
C|Genetics:
A|Map position: 4COP9-4G3845

[illegible]

```

Db      127 PEMQKESVFCIMQJFDCLLTVHERCKKGLSLAKRLNSLAFF---CYDUL----- 176
Qy      1045 QNNNTYSCAPGDPKTLFEYKFEPLRVGNHNYIPLNPMPEFGKRIQRYODLDYSL 1104
Db      177 -----YII-----EPCOVE----- 186
Qy      1105 TDFECNHFVGLLREVEGTALQEPREVALIAISVKNLLIKHSPDDRYSRSHQARIAT 1164
Db      187 -----LGRILIVILLCHGHEPARQAKEGKUYIQA 216
Qy      1165 IYLPFLGLIENVQIRINVDVPPRYNAQMYKDSL--ALPAVNDLVPQKSTIDNSL 1222
Db      217 LYFPFVGQIT-----DEMPVFYNLATEKREVLIGVLOIVRLDSTSLVKMQQSI 267
Qy      1223 -----HKDLGASISGASPYTTSTPNI NSVRNADSGSLISTDSGNS 1264
Db      268 ATRLTYFKIMEBCLLFEHKKAADSLG-----GNSHGPV---SEGAG 308
Qy      1265 IPERNSEKNSLDKIQSSSTLGNVVRCDLQSEIKSLIMCFLYLKS--MSDMLFTYM 1323
Db      309 SPKYSERLSPALNNY--LSEASRQEVRLSEV-----TLISFGGLTKTKLECPITFAY 360
Qy      1324 -----NKASTSELM-----DFTTSEVCLHQFOYMGKRYIARTGMMHARLQOUGSLDN 1371
Db      361 MNLGNAQWPSSESLIQLESTPNGYLMQVNSQLASPSQPSYSREALAQASGRIGASAO 420
Qy      1372 SLTFNHSYGHSDADYHQ--SLEENIATEVCLTALD--TLISFLTAFKQQLADHG---- 1424
Db      421 ALR-----ESLHPILRQKLEMBEENAVTSLSQVLEITENSSMAASHNITDYGKIDC 474
Qy      1425 -----HNPLMKVVDVYLCPLOKHQSETALKNV--FTALRSILYKF 1463
Db      475 ITTILTSFESRNSLAFMKAFPIPIRIFDLHGATLMAENBRFLQOIFHLIRLAVYN 534
Qy      1464 PSTYEGRADMCALCYEILKCCNSKLSIRTEASQLYFLMRNNFDYTGKKSFWYTHLQ 1523
Db      535 DSVKRAVIGLQILVVKHLESCKDSL-----FLT-----IRLAL 570
Qy      1524 VIISQSLADY-----VOIGTRFOQSSIIINNCANSRLIKHTSPSD--- 1568
Db      571 LTTTSELMSDVQVTHMKSNDTLSESGARRLQOISENADEKSYNLLRECGLPDITLL 630
Qy      1569 -----VKDLTKRIITYLMATAQKEHENPMLVLDQ-----YSLAKSY 1607
Db      631 IIPKFTENRMMAEYKHLSDSL--VLADASLGHLLASVAMBRYAABEFYKLGMAF 688
Qy      1608 ASTPELRKTWLDSMARIHVKNGLSEAMCYVHTALVAEYLTTRKAVQWEPPLPHSHS 1667
Db      689 AEPVDLHIMMLHLCDHQEMQSWAQAQAVAVAGVIMQALVARNDGVMS-----KDHV 743
Qy      1668 ACLRSRGVFRQGCCTAFRVIITPNIDEBASMMEDVGMQVHNEDVLMELLQCADGLM- 1726
Db      744 SALRKI-----CP--MVSGETTEASAEVEGYGASKULTVSAVYTL--QLANKLFS 791
Qy      1727 KAERYELADIYKLIPIYKERRDFERLAHLVDTLHRAYSKYTEVHSGRRLL--GTYPV 1785
Db      792 QMBELHFCAISLELYIPVYSKRAYQKACHTLNLNYESLIDGNSNIPITDIATYIV 851
Qy      1786 AFPGGAOYQOTDSETVGEPFEDDEGEYIYKEPKLTPLSISQRLKLYSDKFGSENV 1845
Db      852 GFYGEK-----FGKLDKREYVREPRDVLGDIIMELSHIYESRMDSNMI 896
Qy      1846 -KMIQDSGVNPKDLDKRAYIQTHTVIFPPEBEKLEQKTEPEFSHNIRRMFEWPFQ 1904
Db      897 LHIIPRSQVKAEDLOAGVCYLOITAVDVMEDDEDGSR--ERLFSI-----S 943
Qy      1905 TSKRQGVVEQGRATILTAICFPYVKRIPYVQHTDNLPIEVAIDEMSKVAELAQ 1964
Db      944 TGSVARTGS-----FPALVNRLLVTTSESEJEPSPVENALIMETRTALAN 991
Qy      1965 LCS---SAEVDMI---KLQKLGQSVSVQVNAGPLAVARAFLDNTNTRYPDNKYKLLK 2017

```

```

Db      992 ELEBPRSDGDHLPRLQSLQRILOGSVAQVNSGVLSVCTAFLSGPRTLRLSQELQOLI 1051
Qy      2018 EYVRFVEACGQALAVNERLIKEDQLEYOEEMKANYREMAKXELS 2061
Db      1052 AALLEFMAVCKRAIVHFRLLIGEBDEPHQTQVNGFQSLTABELS 1035

RESULT 5
T18868
Annoblast city protein homolog - Caenorhabditis elegans
C.Species: Caenorhabditis elegans
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C.Accession: T18868; T20461
R.Cummings, P.
Submitted to the EMBL Data Library, October 1996
A.Reference number: Z19034
A.Accession: T18868
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1759 <MT2>
A.Cross-references: EMBL:Z81032; PIDN:CAB02723.1; GSPDB:GN00022; CESP:C02F4.1
A.Experimental source: clone C02F4
R.Wild, A.
Submitted to the EMBL Data Library, October 1996
A.Reference number: Z19278
A.Accession: T20461
A.Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: DNA
A.Residues: 1-1759 <MT2>
A.Cross-references: EMBL:Z81054; PIDN:CAB02887.1; GSPDB:GN00022; CESP:C02F4.1
A.Experimental source: clone F01D4
C.Genetics:
A.Gene: CESP:C02F4.1
A.Map position: 4
A.Introns: 11/1; 130/1; 194/3; 240/2; 313/1; 676/1; 702/2; 746/2; 841/3; 895/3; 1001/2;

Query Match
Score 429.5; DB 2; Length 1759;
Best Local Similarity 18.8%; Pred. No. 3.5e-16;
Matches 389; Conservative 328; Mismatches 758; Indels 589; Gaps 92;

Qy      201 FFHLIQLDGSGYKFFFLKDLQKPKGSIPLG-----FLYGVSP--RNKVRRA 247
Db      7 YVAIVQ-----YTFEDPLPSCDAPRLFLFIDRLCVYQGHGHWGGRKFKDEKNGCGIYP 61
Qy      248 FE-LKMDKSSGYLADSEVMEEMITLN-----KIQLNFEAMQKXRGDSHED----- 298
Db      62 LAYVOIVQKSMFVSTSDYLVADEISRVINEMWTIKEL-----WETTRIGSFEDLMD 116
Qy      299 -DE-----OSKLESGSGGLD-----SYPELAKSARAERIKLSSESRVLFYLDPA 344
Db      117 FNEILLIKTKIESGIPRIEELSKLARVSKLVDRGNTLIGQDVVIRNEGVPL-----DV 171
Qy      345 QKLDSSAEPVKSFEKFGRIILVKNDSLNFNIQCCVAENBEGFTTVBPFVTLSPD 404
Db      172 ESLSL-----LRTYEAHISK-----GRVQSLMRKREPVNTIN--DSFSLLSIKS 215
Qy      405 IKVNRKISADPHVDLNVSVQMTATTSPALM--NGSGPERQSLRGLHLEAMQYPRQG 462
Db      216 VELHCKYSCESISLYLDL--KMFPTTDSYTFPLMNSGSKHTDLNKAFLTPAKEDIQK 274
Qy      463 IFSVTCPPHPIFLVARIKVLQSGIT-----HCAEPYKSSDSKVAQV 507
Db      275 YLMTV-----RVYHVSPISSNATMKHGHHEATIKTYCRGSY--ASDIMMSISIF 325
Qy      508 LKNKAOACQRLGQVRMPAMARTLPEKASNLGNAR--FAAIYQDSNKLSDMDLKL 566
Db      326 LAGVGEAKR-----RVFLNRBEPLPLSLKAYATVNI PKN--LSNDMEFTKL 372
Qy      567 ADFRKEGMAKLPVILGNDLITIDNVSSDPNNYNSYIIPKQPTKCTPITEVEEV 626
Db      373 ISTQ-----LVPGN-----VSOIKARHPLP--SFCPAAILNRADRTAVSIEDSR-- 415
Qy      627 PCIPHTQPTIYTNHLTVYVPEKLYKQDSQKFAKARNIACIEFQSD---EEDSQPLKC 683

```

```

Db 416 -----NMVITLMOAELSGSS-----DNIEARLHVSNNGHVENVEFIS 458
Qy 684 IYRGPGPV/PTRAFAAHLHHNHPFYDEIKETLTQJHEKHHLLTFEHS CDSSSG 743
Db 459 VTGSQSLSTYK-----SIVYHTDKPMTEPIKIALSCASHDYLLILYSSKAYKXP 514
Qy 744 STKRDUVETQVGYSWLPPLKDG-----RVVTSEQHI PVS-----ANLPSGLGYOE-L 791
Db 515 EKGPFALAHVQILRS-SALLCOSEHDLAYKIDNPSHDESVHVMNLPDRRTIKESI 573
Qy 792 GNGRHYGPEIKWVDGKPLIKTSTHVSITYTODQHLNFPQY---CQTESGAQALGNE 848
Db 574 GSAKPHSGOFSLSB--KSFWMISTHSCSSMLTQNEHLNLRVWRCVMLTSLVALAOP 631
Qy 849 LVKYLKSLHAMEGHVWIAFLPTILNQLPVL--TRAQOEAVAVNTEVIIHVAQCHE--- 904
Db 632 I-----GDTEHEMIRFPFSLDLALFEIMHORESEKV---VFDVIAVALRICEPRH 680
Qy 905 -----EGLESHLR--SYVKYAYK-----AEPYVASEYKTVHEBLTKSMITILKPSADFLTS 953
Db 681 YPQAAKIFEVYLKRFSPFSAALKILKCLMHVYISDESDSNEK-ARAAFKVMGSLFCLVYV 739
Qy 954 NKLKYSWFFPDVYLKSMQHLIE--NSKYKLIHQRPASVHH-----AVETVNMMLPH 1007
Db 740 SKKCGIKFEFDPEKTYKOYLREFMKSLVALMSEKKAKMTVQNTALKSIPITIDLL--- 796
Qy 1008 ITQKRPDNPASNAHSLAVFKRCFTMDRGFVKQINNYSCAPDPTLFEYKPE 1067
Db 797 -----NESDSVSEPNLCGFIVDLMNFGSNITVREKLNFLA-----QIVETR- 839
Qy 1068 FLRVGNHBYIYR-LNLPMFPGKRIQRYODQLDLSLTDEFCRNHFLVGLLREVTAL 1126
Db 840 FSLSTICREQLLPCLQIALDI-----IOLDMAT-----EKGEPA 874
Qy 1127 QEBREVLIAISVLKULLIKHSFDRYASRSHOARIATLYLPGLIENQIRNVDVS 1186
Db 875 DBAAECASITIAILERTFLDAKSGEENARELTSFLIMYRPLVQAMIR----- 923
Qy 1187 PFPVNAQMTVQKESLALPAVNPLVTQKSTLDNSLHKDLGASIASPYTSTPNINS 1246
Db 924 -----VIHDKHTD----- 932
Qy 1247 VERNADSRG-----SLISTDSGNSLPERNSEKNSLDKHQOOSTLGNSVVRCDKLDSE 1299
Db 933 ---DARHGFFSVITALLDKMSQMFSEYVEERSSSIDRDLMEVWQMI--RDLNRA 987
Qy 1300 IKSLLMCFYILKMSDDALLFTYWKASTSELMFTTISEVCLHQF-QYMGK--RYIART 1356
Db 988 PFTWMDMIMLQNKVIHKSIL-----RFVMSAVOTFFSNDRKFCVEMWREYMTVVSFVTOE 1042
Qy 1357 GMMHARLQOLGSLDNLFTNHSYGHSDADVLHQSLLEANIATEVCALTADTSLFLARK 1416
Db 1043 G-----LNSKHEWKNEDDMRIQLRKA-----AAQDLASMSRRLTRP 1079
Qy 1417 NQLLADHGNPLMKKVPDVYLCLQHQOSETLKNVFTALRSLYKFPSTFEGRADMQA 1476
Db 1080 SQKLL--NVIIPMISGF-LKVSIVDDEDRERATPIFFDMQREY----- 1120
Qy 1477 ALCYELKCCNSKLSIRTE-ASQLLYFLMRNNFDYTGKGSFYRTHLYQYIISVQLIADV 1535
Db 1121 -----NTSASRSFTEFASSELVSOLDTVNDQHSATKGFKEHFRQSLITLCOGDEL 1170
Qy 1536 VGIIGTRFOQSLIINNCAHDR---LTKHTSFSQDVVDLTKRIRFTVMATPQMKHEHN 1591
Db 1171 MANGBELLERI-----DRLLTALLBYHEVAS--KSLVECDVLSKRVQLMRRYN 1219
Qy 1592 DBEMVLDOYSLAKSYASTPELRKTLWIDEMARIHVNGDLSEAMCYV--HTALVAEYLT 1650
Db 1220 -----QYSHKELVVK-----YIYKLYDLHISYGNKIEAKTKLLHNAHMLTFD--- 1261
Qy 1651 KREAV-QMEBPILLPHSHSACLRRSGVFRQGTAFRITTPNIDEBASMEVDVGMQDVF 1709

```

```

Db 1262 -DDALPQW---LIARS-----LNRHR-----EL 1280
Qy 1710 MEDVAMELLEOCADLMKABERYELADIYKLIIPYER-RRDFEHLAHLDTLHRAYSKV 1768
Db 1281 HQKLEDLMEBAGNLFSGEDBEDALIVYNQVLVYQNIIMDYDLKAGLLOKIAOYISI 1340
Qy 1769 TEVMSGRLLGTYYRVAFFGQAAQOYPTDSETFDVEGFEEDEGKEYIYKPEKTLPLSEI 1828
Db 1341 SRT-----BRAVYYLVAFYQG-----GFPAYLNGHKFVFRSEKLEMHGEF 1382
Qy 1829 SQRLKLYSDKFGSENVKMIQDSKVNPKDLSKATYIYQVHTV-----FFDEKELOE 1882
Db 1383 MORINKMY-----DNPEKIMKTDP-CPLVDSPEKRYIQVFNIDIGTCGSENNPEVKP 1435
Qy 1883 RKEFERSHNIRRFMEHMPQTGKRGQVEGQCRRTILTAHCFP-----YVKKR 1934
Db 1436 YKKTFRYYNIQTEYSK-----VEE--RKTQWTSID--PSEEFKMLVRR 1480
Qy 1935 IPVMYQHHITDL-----NPTEVAIDMSKVAELRQLCSSAE---VDMTKL 1976
Db 1481 IKTADSLPTDLRFTEIVELSDPIYVTPLONAVQEKKNKELNETPAASNPFDKLL 1540
Qy 1977 QLKLOGSVVOYNAGPLAARAFLDITNTKXYPDNKVKLLKEVRFQFVACQALAVNER 2036
Db 1541 SRDILGVVSAAMWG-----VKNY-----EVF--FTBACRNICEGB- 1575
Qy 2037 LIKEDQLEFYQEMKANYREMAKELSEIMHEOI-----CPLEEK--TSVLPS 2081
Db 1576 -----OSVIMELSLIIEVVEILEYCCYHABSCQEBARINMTLPTH 1618
Qy 2082 LHIFNAISGTPSTVWAGMTSSSS 2105
Db 1619 ANI--RSLSPDSDSINDGMWTKS 1640

```

```

RESULT 6
T03284
myoblast city protein - fruit fly (Drosophila melanogaster)
C.Species: Drosophila melanogaster
C.Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 17-Nov-2000
C.Accession: T03284
R.Brickson, M.R.S.; Galletta, B.J.; Abmayr, S.M.
J. Cell Biol. 138, 589-603, 1997
A.Title: Drosophila myoblast city encodes a conserved protein that is essential for myob
A.Reference number: 214881; MUID:97392689; PMID:9245788
A.Accession: T03284
A>Status: preliminary; translated from GB/EMBL/DBJ
A.Molecule type: mRNA
A.Residues: 1-1970 <ERI>
A.Cross-references: EMBL:AF007805; NID:g2367443; PID:g2367444
C.Gene: c18
A.Gene: mbc
A.Cross-references: FlyBase:FBgn015513
A.Map position: 3

```

```

Query Match 3.5%; Score 382; DB 2; Length 1970;
Best Local Similarity 19.0%; Pred. No. 2,3e-13;
Matches 337; Conservative 286; Mismatches 596; Indels 553; Gaps 88;

Qy 500 SSKVAQVYLKNAQACQRLGQY--RMPAMARTL-----FKDASGULD-----KNAR 545
Db 320 SSRKASQLSVSSSGSSNGEYIIRPVGACKDLPINKSDDRGNIDLPIWCEKET 379
Qy 546 FSAIRQOSNLSNDMLKLLADPRKPKMA-KLPIVILGNDITIDVNSDPPNPNVNSY 604
Db 380 LDGTLRK--LIANDIKTI-----DSKMAVTVLEVLKGD---IKQIEEPRLMHTN- 425
Qy 605 IPTQGFCSKTPIFVEBEFVPCIPKATQYTI---YTNHLYV-----YPKYLKXDS 654
Db 426 -----VPAKKGFPFVILPGVNRDLYLTITSGEFAARAK-TS 463
Qy 655 QKSPAKANNIAICIEFKDSDSDSQPLKCIYGRPGPV/PTSAFAVALHHQN--DEFYDE 713

```

```
Db 464 EKOV---EVSVCVA---NEQGYLMPGVLSIGAGHOP---DEYKSVVYVYHDDKPKMOET 513
Qy 714 IKIEIPLQALHEKHNLLTFEFHVSCDNSKSGSTKRDVETQYGVSWLPLLK-DGRVVTSE 772
Db 514 KIHNPIDIEPFOCHLAFVLKH-----KSSNEQKORTEKPRGLAVRLMQANGTITIGG 566
Qy 773 OHI-----PVSANLPBGVLGYOELGNGRHYGPEIKWVGSKP-----809
Db 567 QHILAVYKIDHKKYDKTVANNC-----YLEL-----PATVAELQAKPSIGGLTLLPKD 614
Qy 810 LKISTHVLVSTYTOOHNLHNPQYOCOKTESGAGLGNELVKYKSLHMEGHVMALEP 869
Db 615 QLSIGVNLCSYKLTQSVSLGLLMSAHKST-----LEQSLNLSTVPEEYVVKFIQ 666
Qy 870 TILNOLFVLTATQGE-----VAVNVTYVYIHHVAQCHEEGLESLSRYKAVYKAYBPY 924
Db 667 DILDALFILVENDHEKTDQVPMSTIH-LIETVSDLKHQPLSLVDVYINSEFS-----721
Qy 925 VASEYKTHHEELTKSMITLIKP---SADPLTSNKLKYSWFFEDVLIKMAOHLIENSKY 981
Db 722 FLIAYTKLMDVLOKNISEAISPEKESAD-----GNDLESPEV 759
Qy 982 -KLNRQRPASYNHNAVETVYVNMMLPHITOKFRDNEBASIGNANSLAVITKCFITMDG 1040
Db 760 RELYKTRY---LHYVVKFVIRSRVLY-----AENCNCTYMDPATRLQELL-BK 805
Qy 1041 FVFKQINNYISCFAPGPKTLFEYKFEPLRVVCNHEHYIPLNPMFPKGR-----I 1092
Db 806 FI-----DWIGC-----PSNLKSEGALLK---NHIIATLDMQVBEHVRISISIVEL 851
Qy 1093 QRYODIQLDYS-----LTDFPCR-----N 1111
Db 852 EKFPRLRLTQSKMGKICKDFETGLFTLPKCRALLTVPCFKIHIDHESKEEIAECININW 911
Qy 1112 HFLVGLLREVGTALOEFREVLIAL-SYLNKLLHSGSDDRYASNSQARLATLYLP 1170
Db 912 NTLKLLFRSDVOSTNHDIDIMILFRYTWK---AAHALD-----RTGLGKFF 958
Qy 1171 GLLIENVGRINRVDSPPFNAGMTYKDSLA-----LPAVNPLVTPQKSGT---1217
Db 959 AATMLGLQMDAQHYFYF-----VKDHOSELKHFYIETILVEELVSPHQKAVFR 1011
Qy 1218 --LDNSLHK--LIGAISGIASPYTTSTPNINSVRNADRGSLISTDSGNSLPERNSEKS 1273
Db 1012 DMMDMIMHONTVILGALKHL-----TVVID-----1037
Qy 1274 NSLDKQSGSTLGNSSVVRCDKLDQSEIKSLMCFYILYSMSDDALFTYWNKASTSELD 1333
Db 1038 -----YFLC--PEKQIWSNFPQCSIAFLVQ-----SPLQLND 1068
Qy 1334 FFIISEVCLHOPQYMGKRYIARTGMNARLQ--QLGSLNSLTPNHSYGHSDADY--LHQ 1389
Db 1069 FPDNK---RQIVFAFYRIDIKDTAMEIKMFQLOQ-----HKRFYPOLV 1112
Qy 1390 SLEANATEVCLTALDLSLFLAKNQLADHGNPLMKKVFVYLLCFLO---KHOS 1445
Db 1113 PLEMSMIEKELRQ-ETIPIF-----FDMQGEYYSRRHEHS 1150
Qy 1446 --ETALKAVFTALRSILYKPFYFEGRADMCALCYELKCNKSLSIRTBASQLLYF 1503
Db 1151 YGDT-----KFNNAHHNG-----NFSDFKTAMIEKLDI 1178
Qy 1504 LMRNPDYTKGKSFVATHLOVIVSGLIA-DVVGIGTRFOOSLSIINNANSDDLKH 1562
Db 1179 LIG-----AGKDAEYKHLPETIMLRCAANTLNDGAFVQVMTRL-----MDKLLEY 1228
Qy 1563 TSPSSDVKOLTKRIRTLMATAGMKEHNDPEMLVDLQVSLAKSVAST--PELRTKWLDS 1620
Db 1229 RFLIQD-----ESKENR-----MACTFSLLQFYSEVDLKEMYIAYVWK 1266
Qy 1621 MARIHVQDGLDEAMACY-VHTALVAELVTRKEAVQVBPRLPHSHSACLRRSGQVVR 1679
Db 1267 LCAALHMEFENYTEAFTLKLHT-----ELLRMTDTLSHQ-----LRSYR---H 1307
```

```
Qy 1680 QCTAFRVTLPNIDEASMMEDVGNQDVHFNEDVLMELBQCADGLMKAERYELADIYK 1739
Db 1308 NNCRRHRLQ-----KEALYFE-----IMEXFD-----XGKQMECALDMCR 1342
Qy 1740 LIIPYERK-RDEFLANLVDTLHRAYSK-VTEVMSGRRLIGTYFRYAFFQQAQYOFT 1797
Db 1343 VLARQYEEIFDYKLAEILNRMALFEYKIKELHNSB-----YFRVCFYGR-----1390
Qy 1798 DSEIVGEFPEDEDEKEYIYKPKLTPLEISQRLKLYSDKFGSENVMIQDSKVPK 1857
Db 1391 -----GPRFLQNRVYIFRGEYERHSDFCARLVQHPQ---ALHQTEAPG---D 1436
Qy 1858 DL-DSKAVYIQTVAHYIPF-----FDEKELOERTKEBERSHNRFRFEMP--TQTK 1907
Db 1437 DITNSDQGYIQVKNKTEPIMGAFNFKXINNELVKYFTANNVQKFPSPRFRSTNG 1496
Qy 1908 RQGVYEOCKRRTIITAIHCPYVKRIPVYQHHTDINPLEVALIDMSKVAELRLQC- 1966
Db 1497 DRDDVRNMLERTBELRISYPLRPGILRWFPVETNFKISPLERAVEIMKDTNRDIQLVI 1556
Qy 1967 ---SSAEVDMITLOKLOGSVSVQVNAGPLVAYAF-LDDTNTKRYPDNK--YKLKAYR 2021
Db 1557 LKKSDETHINPLSKMLGIVDPAWGGPAAKYEAPLTDDYUEQNPDPDEBELKELIA 1616
Qy 2022 QFVEACGALAVNERLIKEDOLE-YQEBMKANYREMALESEIMHQICPLE-EKTSYL- 2078
Db 1617 NQIPLDLALQGH-RLRAPDSIKALQEHLERFADPMQGHVEGRYKSGCDLKIENDSVVM 1675
Qy 2079 --PNSL--HIENALSGTPTSTVHGHTSSS 2105
Db 1676 RRPNSFLPSLFDGNSNRHSET---SMGSSDS 1703
```

```
RESULT 7
BS3435
vesicular transport-associated repeat protein Tb-292 - Trypanosoma brucei
N.Alternate names: membrane-associated protein Tb-292
C.Species: Trypanosoma brucei
C.Date: 26-May-1995 #sequence_revision 07-Jul-1995 #text_change 07-May-1999
C.Accession: BS3435; S34395
R.Ilee, M.G.S.; Russell, D.G.; D'Aleandro, P.A.; Van der Ploeg, L.H.T.
J. Biol. Chem. 269, 8408-8415, 1994
A.Title: Identification of membrane-associated proteins in Trypanosoma brucei encoding a
A.Reference number: A53435; MUID:94179225; PMID:8132566
A.Accession: BS3435
A.Status: nucleic acid sequence not shown
A.Molecule type: mRNA
A.Residues: 1-189, 558-721, 'P', 723-738, 'D', 740-829, 'L', 831-1100, 'Q', 1102-1636, 'L', 1638-16418, 'H', 2420-2499, 'L', 2501-2550 <LEB>
A.Cross-references: EMBL:X73956
R.Ilee, M.; Russell, D.; d'Aleandro, P.; van der Ploeg, L.
submitted to the EMBL Data Library, December 1992
A.Description: Identification of membrane associated proteins in Trypanosoma brucei enc
A.Reference number: S34394
A.Accession: S34395
A.Molecule type: mRNA
A.Residues: 1-2550 <LEB>
A.Cross-references: EMBL:X73956; NID:g393395; PID:g393396
C.Keywords: tandem repeat
F.74-557/Region: 8-residue repeats (A-R-L-R-A-E-E-E)

Query Match 2.6%; Score 279.5; DB 2; Length 2550;
Best Local Similarity 17.9%; Pred. No. 2.9e-07;
Matches 375; Conservative 306; Mismatches 758; Indels 655; Gaps 94;

Qy 245 RPAPFLKM-----QDKSSYLLAADSE-VEMEMITLTKILQTLNFAAMQEKRNGDS 295
Db 761 RFFENELMSDPTVTYTBEDRNLPNHSRDEQVLEBQGLRTKRNQNNLV---AALSLRFLTGGP 816
Qy 236 HEDDQSKLBSGSGSLDYVELAKSAREAEIKLSERVKLFLYIDPDQOKLDPSSAEP 355
Db 817 VQVWEQA-----CGFSLDP-VPGRIDGVRLPHTHYVAGSSYIV-----AHRIATNKVPYA 866
```


QY	336	UASFEKRGKRLVYCNPLSFLQCCVAMENEGPPTTNEPFWLTSLEPDIKXNRISA-	413
Db	867	MKAFFS--GVLLRVECTSIASRL-----EVHMDADVPIVSAFSSMSSTKLKXSE	917
QY	414	---DFVLDLHNFVSVRQMIATTSPALMNGSGEPQSAIRGLIH--EAMQYPKQIFSVT	467
Db	918	TFFPFSQVDIFY-----PHKE--RSEINRESRAMAETPOEFFA--	953
QY	468	CPHPDIFVAREKYLQGSITHCAEPYKSSDS-----SKVAQKYLK	509
Db	964	---SLVLMRYR-----PCEBODSYVDLYSRPDKYSQOVTLMKONTOLLA	998
QY	510	NAKQACORLGOYRMPFAMARTLFXDASGNLDKARFSAIYRQDSNKL-----SNDWLK	564
Db	999	MKSDIFEEIGWYMPCV-----KXK--SMVENTLETKLYRGVDPACQFOQ	1041
QY	565	LLADPRKEKKAUKLVLILGNDITIDNVSDDPYNVSSYIPTKO---FETC--SKPIT	619
Db	1042	LVENDGAOKALATLTFPFAFSLKHKCGEVEFPQ-HDEPPEENETIIKCTRSDJEVQ	1100
QY	620	FEVEFEVP-CIPKHPQPTIYTNHLXYVPKLYKQSOQSPAKA-----RNIACTE	669
Db	1101	BEHCFFVSIIPILNSGF--FNSYNNV--YFRDRLVLTSGILATVPSTHRRFFWMEIN	1156
QY	670	FKDSDP--DSOPLKCIYQ--RPGGPVFTSAPFAVALHHQNPBEFYDEIKIELPTQLEKHN	727
Db	1157	VRDNTSLSGENTPLLYGNRLSAKTLQTSVMAVAHNSLDPSLSBFPQOLPLHIGKEFH	1216
QY	728	LLLFEPHNSC---DNSSKGSFKKRDVETQYQWMLFLNDG-----RVYSEQ	773
Db	1217	IFITLY--ACCKKDNSSBENQKH-----KGTAFPIINHGVVRKDEMTIKFIADQ	1269
QY	774	HIPVANSPLPSGYL-YOELGMRHRYPEIKWVGSKPLKISTHLVSTVYQDOHL----	828
Db	1270	---VLEGSEKSYLDFSEADBAH-----LNGDIPVLSYSTQRTSVHASMNIASL	1319
QY	829	-----HNFQYCOKTE---SGAOLGNELVKYLKSLHAMEGVMIAFLP--	869
Db	1330	KEMPASLESILRNDLFRVSGNLKDIYRSDDAIHSSILIRKRLPLAE---IIAFYPL	1376
QY	870	-----TLNQLFVLFRATQOEVA-----VNTTR	893
Db	1377	SFFILSLSSPSKSVSLPCTRTALEVLDDITIKQHYDLTTRSKQKQHSAGVSKTSTVR	1436
QY	894	VIIHVAOCHBEGLESILRSYV-----KXVAKABYV-----ASEYK	930
Db	1437	FLYH-----HLTNDLILYNGEKQRLYAGFAETMLILVACRPHSENTQAGDK	1484
QY	931	TYHELSLTSMTTLKPSADFLSNKLKXSPFPYVLKSMAGHLENSKY--KLRLNQ	987
Db	1485	TAKDREDDKKNIRKGMADL-----SWLEFVILIRSYIYMALEPSPILPRAKLLD--	1533
QY	988	RFPASYNHAEVTVMMLPHITQKFERDNPBASKNANHSIAVFIKSCFPMDFGEVFKQIN	1047
Db	1534	--PGRYSIIIGLCYBAL--DVLDEGIDDLVLRRT---ALFTNRLFPCORGVANTIQ	1585
QY	1048	NYISCFAGDEKTELFYKFEPLRVVCHENHTPLNLPMPFGKRIQRYQDLQDLSLDE	1107
Db	1586	RIVERF---EKQDMEGIGTFMSIILDVDVYTLMI-P-----SSS	1622
QY	1108	FCRHNFLVGLLREVGTAQOE--FREVRLIAISVKNLLIKSPDRYASRSHQARIATLY	1166
Db	1623	YTOPFELIRILVHAFSLSSPSQVBRSSSDVLKYPICRTYNGVRYP-----ATCL	1674
QY	1167	LPLFGLLIENQRIINVRDVSPPPVNAGMTVXDESIALPANNPLVTPQKGSTLDSNLHKDL	1226
Db	1675	RWMAQSQALALVREVSIR--WKAVQVQCKTESIVA-----TIEDK--ROL	1714
QY	1227	LGATISGI---ASPYTTSPTNINSVAMDSRQSLISTDSGNLPLPRNSBK--SNSDKIQOS	1282
Db	1715	CVSILMITIYVAP-----GALRWLKG---EDSDSVISIGLLSLVSDA	1753

QY	1283	STLNGSVRCCKLQOSEKSLJLMCFLYLTKSMSDALFTYWNKXSTSELMDFPTI--SEV	1340
Db	1754	QHLRYRGAQVJXANHGVKEP-----TQMRMRDARMESTFVAIGAVY	1796
QY	1341	CLHOFQYWKRYIARTGMAHRLQOLGSLD---NSLTPNHSYGHSDADVLHOSLEANI	1396
Db	1797	C-----SAVLHAPDPAKSLRTEKVDPIVF--PPFHLLEGVHLGNSTPAL	1840
QY	1397	ATEVCLTADTLSLFTLIAFKNQGLADBGHNDLMKVVDVLYCPLQKHQSEFALKNVFTAL	1456
Db	1841	QIGSC-----ALLEVAHS-----IRPEILCSNR-----	1864
QY	1457	RSLLYKEPSTFEGRADMCALCYEILKCCNSKLSIRTEASOLLYPMRNPDYTKKS	1516
Db	1865	-----MCSGMAYLWMSLMSCSVHVRMAAEVL-FLM-SYACTGCRS	1905
QY	1517	-----FARTHQVLIISVQLIADVDYIGGTRFOOS--	1546
Db	1906	VAKIRAPISATVYVAESKKRDLRLAGFLEHLISNIVRAAQOSGD---NFPFSGSYV	1961
QY	1547	-----LSIINNCAUSDRIKATSFSSDVKDTYKR	1575
Db	1962	KRCADSETPGNEVYTAGRRHGERPPFSVERMSPIBSCLSINGV-----NGIASLVGN	2014
QY	1576	IRT-VLMAATQMKENHP---EMLYDLOYSLAKSYASTPELRKT-----	1616
Db	2015	RRTSAIVGSEOGSNUNCNEPPSFABEILASVEISMMLEFGSVQLOMTESLRLKAKALQHF	2074
QY	1617	-----WDSMARIHYNQGLSPAAUCYVHVTLV-----AEYLTR-KE	1653
Db	2075	KLAQQLRVGNHECCLRMFORHELHANNNGEAGIVLRFVALCRRLTEAFQVARGKE	2134
QY	1654	AVQWEPPLLPKSHSACLRRSR-----GGVFROGCTAFRVITPNIDEASMBEDVG	1703
Db	2135	SRGARIPPIVPSHFVMDHYARLLPEADTLTLTGVLVTVSSLRVL-DS-DPSLSLSGHV-	2191
QY	1704	MODVHNEEDVLMELLECADGLMKAERYELIA-DIYKLIIRIYKRRDPERLALYXDTIH	1762
Db	2192	-----EVLKQASLLDQ-DHYNIIFSQVLSIVERIYKLLNDFKAAASIHSA	2238
QY	1763	---RAYSKVTEVMHSGRLLGTGYPRVAFPGOAOYQFTDSESTVEGFEDEGKE-----	1814
Db	2239	DGCYAVSREDDQKRENRYFLMAR-----NRTEVE--LHENDGHSNONG	2283
QY	1815	-----YIYKEPKLTPJSEISQ-----RLKLKYSKFGSENVKMIQDSKVNPKUL-D	1860
Db	2284	YSLPLKCFKVPKATKLEOFLCEGSKDFVRPLKATPLVYTEMAEOKDMGALASXLPR	2343
QY	1861	SKAVYIOVTHYIIPFDEKELOERKTEPERSNIRRFEMFEPQTGKROGVEQ-----C	1916
Db	2344	ENCCLSVCEVNPFFSSGR-RKLTDSYDRNASTLKEFY-MTTYDNSREVEGDAPDSAL	2401
QY	1917	KRRTILTAINH---CFPYVKRIRIPVYOHHTDNLNIEVALIDEMSKVAELRQLCGSSAEVDM	1973
Db	2402	RNRLVVVKYHHERSFSTNVIVDAATHIJDOLDPAATVAHVNLKSGIEMLQ-----APVDN	2456
QY	1974	IKLOLTKQGSVSUVNAGPLAYARAFIDDTNTRKYRPNKYKLKKEVERQVPEAC	2027
Db	2457	DELVGTLLRAATPGGFARPGAYIKEVITVWTSYNSAVMTQVRSLSKLLKTKLTKC	2510

RESULT 8
S53409
Probable membrane protein YLR422w - yeast (*Saccharomyces cerevisiae*)
N|Alternate names: hypothetical protein L5576.7
C|Species: *Saccharomyces cerevisiae*
C|Date: 05-May-1995 #sequence_revision 01-Sep-1995 #text_change 19-Apr-2002
C|Accession: S53409
R|Favell, A.
submitted to the EMBL Data Library, February 1995
A|Description: The sequence of *S. cerevisiae* cosmid 9576.
A|Reference number: S53409
A|Accession: S53409

1473 -----ERKEYLKESARNFSGQKPEKALAVYKDLKAYDEINYDNGLAFAVDQIAGIY 1527
1766 SKYTEVHWSGRLLGTFRVAFPGQAAQYQFTDSEFDVSGFDEDECKEYIEPRLTLP 1825
1528 TRLOST-----DRLVPTTFKVSFMG-----FGFPKSLANKSFVEBGLPEPEHI 1569
1826 SEISORLKLKYSK-----FGSENVKMIODSGKVPKDLDSKYAYIOYTHVIP--PFDEKE 1879
1570 TSMHDLRLRSYHSGSNIVHSQEEVDML-----LKNP-----PMGKIYHVASVEPCISIDNYN 1621
1880 LQKRTKEFERSHRIKRFME--MPTQYGRKOG--GVEQCKRRTILAIHCFPIYKGR 1934
1622 SSDKSSSI--NNVRYVYIENRDLRTFSNSRRLPGAGVTDLWVEEYTHMTNFTPLMR 1679
1935 IPWYOHHTDLPNIEVAIDEMSKVAE--LRQLCSSAEVD-----MIKLQKQSGSY 1984
1680 SEIVKTKSLSPLENAIRSLQYKIOBELVLEMCNKTKLDHGDVNDLFTLESTNITGTI 1739
1985 SVQVNAAPLAYARAFIDTNTKRYPDNKVYLLKEVRFQVACGQALAVNERLIKEDOLE 2044
1740 SAVVNGISQY-KAFLEPSTSKQFSTDDLGRLLAFDELVAVLGRCULTLHABELPEKDLK 1798
2045 -----YQEMKANYREMAKELSEI----- 2063
1799 PSHDLLVRLEEFENFAEIERYSRTLSBANRSNNMITARIISHKNPKKASFSGRDHHS 1858
2064 --MHEQICPLEEKTSLVPLNSLHIFNAISGPTSTMYHGMTSS 2104
1859 GSNHSQFV-LEHSDSFGPNSL-LFGKYL--TTTLSHSTTSS 1896

RESULT 9

A90551
conserved hypothetical protein MYPU_3130 [imported] - Mycoplasma pulmonis (strain UAB CT
C.Species: Mycoplasma pulmonis
C.Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001
C.Accession: A90551
R.Chandaud, I.; Helliig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer, I.;
Nucleic Acids Res. 29, 2145-2153, 2001
A.Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm
A.Reference number: A9512; MUID:21267165; PMID:11353084
A.Accession: A90551
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-2619 <CUR>
A.Cross-references: GB:AL445566; PID:914089727; PIDN:CA013486.1; GSPDS:GN00153
C.Experimental source: strain UAB CTIP
C.Genetic: 3
A:gene: MYPU_3130
A:genetic code: SGC3

Query Match 2.1%; Score 225.5; DB 2; Length 2619;
Best Local Similarity 16.8%; Pred. No. 0.00045;
Matches 411; Conservative 399; Mismatches 851; Indels 781; Gaps 106;

15 KFTRA-LSKPGTAELRQSVSEVVRGSLVLA--PKLIEPDEYENVVOKKTOIANDCLRE 72
621 KFTGATQKILSHNSELVLRIVSGEKTFALEKIDDI-PDNKAVYQKVEKIED----- 675
73 MLFPVDDPOTAILRQRYICSTVPAKABEBAQSLPTECIKTYSMDHLVNYKEDYS 132
676 -FLFAVNTNEKAITPFVFLNELVAKPE-----NFKALVSP 712
133 G--EFPQDLPNKV-----KLQDLPHVVEVDEVDK----- 161
713 GTSQFLNSNEVISTQSAFYEVSKTLVAVDTBELKVLKNPIEIKSSAQKLNSP 772
162 -----DEDAASIGOKGKITKHGMLYKGNMNSAISVTRWSFKRR--PHLIQAGDSYK 214
773 KERNLEKQKQKSIISIFGTNDLGMFLKQVYKSKLSSVEQDKTSSLEKVSFVFS 832
215 EFLK-----DLQKEPKGISPLG-FLYGVSPFN-----NKVRRF----- 246

833 ENVKKLLISSAGTYLSNIELFKE-NSDVFKGIFSHILNFEKTEPELLNKTIIDVINSWESI 891
247 -----APE-----LKMQKSSYLLAADESEVMEEMITI-LKKIL-QLNFE--- 284
892 DOYNDAFEFVLNFKLTNSWFTYKDIFFSVVNSEIQANILITSKNLKLSQAGFELTN 951
285 AAMQEKRGDSHEDDEOSKLESGSLDSYLPKLKASAREAIKLL--SESRVYKFLV 340
952 SSVESIKNIFTN-----TKTIADSNITDALVDTFVNSKTLLEIDKKNFLASIKNIQIFAL 1007
341 DPD-----AQKLDSSAPPEVKSPEKKGKLI-----L 368
1008 SDDNFSVFKALISSNKLKDKAFVNOYKODITNVLKEL-LFEKEFPNLIYSPMKTSBL 1066
369 VKNQDLSFNLOCCVANEENEGPTNVEPFFVLTSLFDIKYNRKISADPFVNDLHFEVRQMI 428
1067 FENNEBALNKTTIIRN--ALKKNDPPFIN-KMMDIIFD--NLDKYQKLENYG--QVL 1118
429 ATTSBALMGSGPETQALRGILHEAMQYPRKOGIFSVTCBPDIFLVARIEKVLQSGIT 488
1119 GLIKNSDALATKYOALFKOVVNEANFQDVLK-----VLVVSQTKKFN-GPEL 1166
489 HCAEPYMKSSDSSKYAOKVYLN--AK-----OACQL----- 518
1167 NADETNKISTDLTKIKQVFESNFKSLALATKDSLVIDGVQANKKIFDPTPLNTFL 1226
519 -----GOYMPAMARLTFKDASGNLDKNARFSAIYRQDSYKLNDDML-KLADF 569
1227 SEKDEKENTQMVMS-ILNTKLIDISONSSEIKRLKTFKSA--LSRDEYEKTIKVF 1283
570 RKPERAK--LPV-----ILGN-----LDITDVSDEPNY--VNS----- 602
1284 LSKEKLASVYFVEYFVKLNKTKTNSDILLIDSLPKDFSQYQANSFEELIYVLRN 1343
603 -----STPTKQFETCKPTTFVEEYFVPI--PKTKQPI-----TIYNHLYV 645
1344 NKEIVNVYKELSEFIKSNITIQRLGQTINALVTPNEKIFSDTELQSLSLINESDL 1403
646 YPK-----YLKVDQKSPAKANINICIEFKQSDDEDS--QPLKCIYGRP 688
1404 PEKTIINISLNDVLVSLAINOESSVLTKEVNT--HKAIDQNSYFAVIKILYSK 1459
689 GGPVETRSAPAAVLHHQNPFEYDEIKLEPLOTLEKHLLLTPE-----HVSQDSS 741
1460 EGDVDKSDBFKAIKKPLN-----LEAKDVIYLTKISPIVDKTIQDFDST 1505
742 KOSTKRDVETQVGYSLPLKQGRVYTSQHIYFVSANL-----PSGILGYELGNG 794
1506 KSIYK-----LIKQ--VLSDRONLKFSAQIITDTIFANKKTYLAVENTLN-- 1547
795 RHYPGEI--KWVDGKPELTKISTHLVSTVYTOPQHLNFRQYCOKT--ESGAQALGNEL 849
1548 -----DFLIDALVDNKENIINILSNITLGKITDEPKKIISFRTKNMTSSSTTBEI 1602
850 VKYLSLHAMEGHVIA-FLPTIQLPFRVLTFRATQEEVAVNVTVIIVVAQCHEGLE 908
1603 NTIVASAEHLISIVDSNFIQLKDPALKEKKNKQKSKTSLE-----FS 1648
909 SHLRKYKAYKAEYVASEYKTV-----HEELTYSMTTLKPSADFLTSNKL 957
1649 TFMKKSISEPLDAENNYALFRKYLSSNLSTOSDQSFYSISILIRV--FLRKEKAL 1705
958 KYSWFFPVLLKSMAGHLIENSKVYLLNQSPASVYHNAVEVVMMLPHITQKRDPE 1017
1706 -----ELMLK-----KQFNKPEL--HDIKVDALNL--IKFIFRDN-- 1737
1018 ASKNANHSIAVFIKCTFMDRGFVFKQINNYISCFAPGDKTLEFYKFEELRVACNHEH 1077
1738 -----QVIDPI-----ETIIN-----RVVSQKEE 1756
1078 YIPLNL-----PMFGKRIQRYQDLDYSLTDEFCNNHFLVGLLAEV 1122

Db 1757 YKLSNYSOEILYKPLSSNKNETLNFPKQIGETKNSKLINPLAKSFISEB-----LKKS 1810
Qy 1123 GTAQEFREVRILAI SVLKLLIKHSFD--DRYASRSHQA3IATVLYPLFGLLIENVORI 1180
Db 1811 KTLTQN-DEIDSVVFNVESLVLEKTDLID-----KLYDLVSSFPNAKVSPTKV 1860
Qy 1181 NVRDVSPF--PVNAGTIVKESLA---LPVAV-PLVTPQKSGSTLDSNLSHKDLGAIISI 1233
Db 1861 EIFNFSLSLKPI---MDQKNYFSLVKTATLITNMKLMESQNTDALKNALGOMLBRVFK-- 1915
Qy 1234 ASPYTTSTPNINSVNRADSRGSLISTDSGNSLPERNSEKNSLDKHQOSTLGNSVVRCD 1293
Db 1916 -----NTDLLNKPLDPVLQNIINPLDQOKSKTKENLRA-- 1949
Qy 1294 KLDQSEIKSLMCFLY-----ILKMSDDALFTYNNKASELMDPFTISEVCLHQFOY 1347
Db 1950 ---NASKQILVFLVLONEOTLEPFKSLVNNIISNNQNYGANNVYDLIY---SFIENKQ 2002
Qy 1348 MGRYIARTGMMHARLOQSGSLDNLSTFNHSGHDADVIHOSLEANIATEVCLTALDT 1407
Db 2003 LVTOYFKKIAPKIASDSMLKTFVNGFALNLOEQN---LELSKDELSVSNITSGAIDA 2058
Qy 1408 LSLFLTAPKQULADHGNPLMKKVVDVYLCLFQKHQSEITALKNVVTAIRSLIYKPEPTF 1467
Db 2059 LST-----SNIVEKILEVFLAEVKKQNPQ3KTNDDLPA3IQ----- 2093
Qy 1466 YEGRADMCALCEYLKCCNSKLSIRTEASOLLYPLMENPDYTKSGSFVRTHLOVITS 1527
Db 2094 -----ELIK--SKFDSQNGLFEVLSIILKSN----- 2118
Qy 1528 VSQIADVVGIGTRPOQSLIINNCSNDRILKHTSFSSDVK-----LTKRI----- 1576
Db 2119 ---LDSKSNINKEYIQVKYTL-----PEKLIKNTSVQNFYDKKTIENISISPIEQNN 2169
Qy 1577 ---RTVLMATQMKHEHNDPEMLVDLOYSLAKSVASTPELKK-----TWL 1618
Db 2170 KQSLSKIVALLASDPOTTLVNVKVIDIDIFKNHKNVASKANLEIILAEPLTHKNOQLAYL 2229
Qy 1619 DSMA-RIHVNGDLSAAMCVVHTALVAEYLTR-----KEAVQWE----- 1658
Db 2230 KATINKLVKSELHDFS-----VNVLLSQLSKKNIEFNNEKEQKSLKTSIHLIIN 2284
Qy 1659 ---PLLPHSHSACLRSSRGVFRQCTAFRVITPNIIDEASME-----DVGQDVA 1708
Db 2285 SDIPQVANNLJBELNKKK--ALSRSVSNQKITRNLNFDKGFBEFKVLANSVQKQNN 2342
Qy 1709 FNEVDLMELLEQCADGLMAERYELADIYKLIPIYERKROBERLAHYDTL--HRAYS 1766
Db 2343 FROEVIINQIKNGSLVNLKDTLINALFTNFDLKIPNDLKQNSFNLKEISNDVHFHRYE 2402
Qy 1767 KTEVWHSGRRLGTYFRVAFPGQAQOFTDSETDVBGFPEDEQKEYIYKPKLTPIS 1826
Db 2403 AILEVAVFSQ-----IHNNGKIYSAQRKSEIKQW--EDLIRVLKKSQSVANN 2450
Qy 1827 EISORLLKLYSDFGSEBNYKMIODSGVNPKDLDSKYAVIQVTHVLPFEDEKELQERKE 1886
Db 2451 KYEDALYDLI-----YKLLKSNLSK-----NILVSQLPMF-----XSAKD 2488
Qy 1887 FERSNIRFRFMEBMPFTQTKROGVVEQCKRRTILTAIHCPYVKRIIPVYQHHTDIN 1946
Db 2489 IYTKHHEILIRFVOGIGRAHEQG----- 2512
Qy 1947 PLEVAIDEMSKYVABEROLCSAEVDMITLOKLQSGSVGVNAGPLAVARAFLDOTNK 2006
Db 2513 ---MLKNIFFKSLDFKNTNFSKVPFNKLTLEIMNIYSN-----FMDKXTNT 2557
Qy 2007 RYEDN--KVK-LLEKVFQFVEACGALAVNERLIKELOL-----EYQEMKANRYENA 2057
Db 2558 LVVSNIMTKDLDLKKYFRTYRDAKQFVNFPMNLFFASLSLPMKTMHQYQ-----NGIY 2610
Qy 2058 KSLSEIMHGOICPLEKTSVLPNSLIHFNALISCTPSTWVG 2099
Db 2611 KMLTNTIY-----PNEANAGA-QGSSSLKVDG 2638

RESULT 10
T08611
hypothetical protein DocA - slime mold (Dictyostelium discoideum)
C/Species: Dictyostelium discoideum
C/Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C/Accession: T08611
R/Auby, L.; Firtel, R.A.; Iranfar, N.
submitted to the EMBL Data Library, August 1997
A/Reference number: Z16456
A/Accession: T08611
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-831 <AUB>
A/Cross-references: EMBL:AF020409; NID:g2425146; PID:g2425147
A/Experimental source: strain AX4
A/Genes: docA

Query Match 2.1%; Score 224.5; DB 2; Length 831;
Best Local Similarity 17.4%; Pred No. 7.3e-05;
Matches 158; Conservative 143; Mismatches 297; Indels 311; Gaps 35;

Qy 1234 ASPYTTSTPNINSVNRADSRGSLISTDSGNSLPERNSEKNSLDKHQOSTLGNSVVRCD 1293
Db 140 AATATATSTLTNNNNNSNNNSITITNNNN--NNNNNNNNNDIGSSAVTGGSGNSLI--- 195
Qy 1294 KLDQSEIKSLMCFLYIILKMSDDALFTYNNKASTSELMDFPTISEVCLH---OFQYM- 1348
Db 196 --NQSQFE-----FTQWR-----SFFLLTSYVINCQDLOFESVN 227
Qy 1349 -GKRYIARTGMMHARLOQ-----LG-----SLDNLSPNHSY 1379
Db 228 PAKAVFITRCGDVAIEARPERVWATVPYKERSTFVLYGLPIYKASISPTMAKRYA 287
Qy 1380 GHSDDVLIHOSLEANIATEVCLTALDPLSLFTLAKQULADHGNPLMKKVDPVYLCLF 1439
Db 288 TKIFPDMLESEILQNGTDLFYHTIDSL---LEISTERL-NRGGWPIVCRSVGDKLPF 342
Qy 1440 LQKHQSEITALKNVFALNSLIYKFPSTYEGRADMCALCEYLKCCNSKLSIRTEASQ 1499
Db 343 -----TLKSP-----STFENVK--CWSL-----VLKSTDAITKYSQ 373
Qy 1500 LLYFLMRNPDYTKGKSFVRTHLOVISVQIADVVGIGTRFQGSLSIINNCSNDRIL 1559
Db 374 FTHDI--NHFLY-----L 384
Qy 1560 IGHTSFSSDVK--DLTKRIRFVLMATQMKHEHNDPEMLVDLOYSIAKSVASTPELAKTW 1617
Db 385 I--PSPNPDVSGDEBEIFFSVSKLLYYLQEHKRTAHFI-----RF 423
Qy 1618 LDSMARIVKNGDLSAAMCVVHTALVAEYITRKAQVQMEBPLLPHSHSACLRSSRGV 1677
Db 424 VMTSRRRHYEMGNYIEAA-----VTLMHLASL--YQWD----- 454
Qy 1678 FROGCTAFRVITPNIIDEASNMEDVGMQDVHFNEDVLMELLEQCADGLMAERYELADI 1737
Db 455 ---ANKVIAQTAASEYSGSPEQK-----ESERKEL--L 481
Qy 1738 YKLIPIYERKDRFERLAHLY-DTLHR-----AVSKTEVWHSGR 1776
Db 482 YKEILLCTNNGAMKRALPIYKELVHNHTNMIQDMSAATYLRQOGAFYQKINE---SND 538
Qy 1777 RLLGTYFRVAFPGQAQOFTDSETDVBGFPEDEQKEYIYKPKLTPISQRLKLY 1836
Db 539 PVFEDYFRVGYG-----KGFSSIONKEFIYKGNQFRLSDPISKI----- 580
Qy 1837 SDKFSSEBNYKMIODSGVNPKDLDSKYAVIQVTHI-----PFPDEKELQ 1881
Db 581 QKMPKSELKTTVEPSQIQDSQOYLLITSVNSNIGFQKRHDSANASITNKGRVP 640
Qy 1882 ERKTEFERSHNIRFMEBMPFTQG--KROGVVEQCKRRTILTAIHCPYVKRIIPV 1938

Db 641 HRVQCFNARNKVVVYVYSKPFKNATATYSQNEFEDLWMNLYFCVGENSFPCTERCLIV 700
Qy 1939 YQHTDNLNPEVADIDMSKVAELROLCSGABVD-----MITQLKQSVSVQVN 1989
Db 701 ERKQVNSIPLENLNSIVQKDEL-----AARIDRHOANTOESISPLTMLNLTIDASVN 755
Qy 1990 AGFLAYARAFLLDPTNRYPD--NKVYLKXEVPRQFVEACQALAVNERLIKEDQLEYO 2047
Db 756 GGVSR-ETFMSEDYIKQNPVEYKNIAELKSAIDQQLVIVEQGLKHSQLRPEMAAMD 814
Qy 2048 ENKANTREM 2056
Db 815 KLETFEITM 823
RESULT 11
G70163
Hypothetical protein B80512 - Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C:Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 08-Oct-1999
C:Accession: G70163
R:Fraser, C.M.; Caajiang, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White
son, D.; Peterson, J.; Keriavege, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 390, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; PMID:98065943; PMID:9403685
A:Accession: G70163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2166 <L2>
A:Cross-references: GB:AE001153; GB:AE000783; NID:G2688419; PION:AA06876.1; PID:G268842
A:Experimental source: strain B31
Query Match 2.0%; Score 222; DB 2; Length 2166;
Best Local Similarity 17.4%; Pred. No. 0.00047;
Matches 407; Conservative 378; Mismatches 800; Indels 752; Gaps 105;
Qy 15 KPTRALSKGTAAELQSVSEVVRGSLAKPKLIPLEYENVIVQKTIQLNDCLREML 74
Db 26 KGSRALDK-----IKKPVDTLTKVLEL-----FIEDKKEINDLAVDM- 63
Qy 75 LPEYDFOFALIRGRGYICSTVPAAKEEAQSLFVTECIKTYSNDMHLVNYKYEDYSGE 134
Db 64 ----EAYQSSIT-----ITKKIE-----VQKIKKNSNPAVEYKKIAHDSM 104
Qy 135 FROLPNKVVK-----LDKLPHYVEVDEVDKEDDAASLSOKGKITKGM 180
Db 105 LKDLDEMTFKVQENIQRLQVQDKIVDKLSKTLKGFTQIDSYE--SNLSV--LEKFDK 159
Qy 181 LKGNMNSAIVTMRFKRFHLLIQGOSYFEEFLKDLQKPKKSTFLGFLYGVSPFN 240
Db 160 ANKENMESIKIASWEKFDNIKELV-----FKID--NLNKEIS-----LYEYDNLN 203
Qy 241 NKVRPAFELKQDK-----SSYLLADSEV-----EMEWTILN--KIQLNFEAM 287
Db 204 IERKRDILVKNKEDLDESPLEKVEFNIGKYSIESSNFYFNKTKLIENTSLIM 263
Qy 288 QEKRNDSHEDD-----EQSKLEGSGGLDYLPELAKSAREALIKLESRYKLFYLDP 342
Db 264 ESKVKNINEKEDFILNRNLBELQNKFKDILVVD-----RSKEIKDLKEDKLVL----- 313
Qy 343 DAOKLDSSABPVPKFEKFGKRIIVKCNDSLFINQCCVAENBEGPTTNVEPFF----- 397
Db 314 -----VDNEISSSSSFKDNVYSRINSLEESKRIEKGKYE-----QVDDVDFKFS 360
Qy 398 -VTLSLFDI--KNRKTISADPHVDLNFHSVGMIAATSPALMNGSPETQSAL----- 447
Db 361 QVBLNKNIYEDYEDKIS--QVDNN--IREKV-ELSLDLNLSKMSVQSGAIDPIKEL 413
Qy 448 ----RGILHEAMQVPEKQ-GIFSVCPRPDIPLVARIKVLQGSITHCAPYMKSSDSK 502

Db 414 EDDSGIYILEFGKRGADIEVSESFKGDINDQLKQLESQLLDVDSNIOEKLIKLDN-- 471
Qy 503 VAQKVLKNAKQACQGLQGRMFPMAARTLFPDASGINDK--NAFSAIYRQDSKLSND 560
Db 472 ----LISFEEINGR-----FNNYSNLNDINAKYALFE----- 503
Qy 561 DMLKLLADFRKBEKNAKLPLVILGNIDITIDVSSDPFVNVSSY----- 604
Db 504 -----GLDSSSKFEENQMEKYSFTDKLTLAGMDERSL 536
Qy 605 IFTQPEFSKTPITPEVEBEVPCIPKATQPTIYNHLVYPRKYLKDSQSKFAK---- 660
Db 537 MYGEKFETLSQ-EATNNYQEFQDLNKLKLENIESFYNNFEKTOETLKYDFNTSLINKE 595
Qy 661 -ARNALICEPKDSDEDSQPLKCYGRPGGVFRSAFAVLHHNHPPEYDEIKELP 719
Db 596 IGRNI--VEFRD-----RYDEYNI-FV 615
Qy 720 TQLHEKHLILTFPVSQDNSSKSTKRDVETQV-----YSMLPLKDGKRVVTS- 772
Db 616 TQL-BESKLOYSKMGEMDSNLKN-----IESQINTNEPFLSLIQDKXGIELSES 667
Qy 773 -----QHI PVSANLPSCYLGYQELGMRHRYGPEIKWVGKPLIKI--STHLVST--- 820
Db 668 VFNDSLH-----QKKAIDHSGMKDELIALNKSLLIDIKVSEELSSATL 714
Qy 821 -VYTDQHLNPFQYCOQKTESQAQALGNEILVYKLSLAMEHVNIAFLPTLINDLFRVL 879
Db 715 KESLEKQVNDMEVLLKKTGDIESLVEKKEKLDKMSYSQSDAALIGKEFINQTEIT 774
Qy 880 TRATQEEVAV-----NVTREIIVHVAQCHE-----GLESHRSYV 915
Db 775 -----KDSVPELDELNNKFPDKNNFVLSKIECDYKLDKPISEEDILNNKSDLNBERI 829
Qy 916 KYAYAEPPVASE-YKTVHEELTKSMITIL--KPSADPLTSNKLKYSWPFEDVLKSM 972
Db 830 ESKLQIVSNIKSDNQIDFLDRISKDILNRKOSINNEVSKLSQSDOSKLINEITVK--I 887
Qy 973 QHLIENSQVYLNRNFRPASHYHAEVTVNMLPHITQKFRNRPASNAHSL-AVFIK 1031
Db 888 ENLLSSGKVDL-----DLIDSEVTTKIK- --LPSIESLSYLYE 925
Qy 1032 RCFETMDRGFVFKQINNYSICFAPGPKTLFEYKEEFLRVVNCNHEYIPLNLPFGKGR 1091
Db 926 KIDFERNGAI-----YSBELLQDINM-----FNKET 953
Qy 1092 IQRYODLDVSL-----TDFCRNHLVGLIREVGAQIEFR--EVLIAISYVKLL 1144
Db 954 RELEENLSKKEFAVLNNGSEEFYKE--VDSLQDKRTDIAFGQANIDITLSLANKFNDI 1010
Qy 1145 IK-----HSFDDRASR-----SHQ-----ARATLVLPLFGLLIEVQVQIN 1181
Db 1011 NKEINGKYNVISNYRGYSENNISKLENEIMHEINLSRRLTRIDSLSKGDENLQK 1070
Qy 1182 VR-DVSPFPVNA-GMTVXD-----ESLALPAVPLVTP-----QKSGTLDNS 1221
Db 1071 ESFDVSKYQVEKELKVDLDDGEA--KINKLVKIEQYKRLBEALDYRTIND 1126
Qy 1222 L--HKOLLGAI-----SGIASPYTT-----STPINSVR 1248
Db 1127 IMQAKERFGEITNELNKNIESKSEFLNDLYKERFYLIESNFEERYSTFLIESEGASIKR 1186
Qy 1249 NADSRGSLISTDSGSLPERNSEKSNLSDKHQSSITLGNVYRCMDKQOSEIKSLMCEL 1308
Db 1187 D-EIYKTLTSD--ENLDIKISEMDQNEIIEORS--KDLIEFEKLODKTKD--CYG 1237
Qy 1309 YLKSMSDALPTYNKAQSTSE-LMDPTISEVCLHOFQYMGKRYAR-TGMHARLOOL 1366
Db 1238 PINSQGE-----IKAGVEENIGNH--DVCIAKYVTLIDDIIVKENEIHKRISL 1287
Qy 1367 GSLDNL-----TFNHSYGHSDADVLHQS-LANAIATEVCLTAL 1405

Db 1288 KSIESTPPSIEKLNLDKVGSCIDKIANDFNLKYIELEERCNQGLNENKIDNKI--KAI 1345
 Qy 1406 DTLSFTLLAFKQQLADHGNHPLMKKVPVLYLFLQKHOSFTALKVFTALSLIYKFPS 1465
 Db 1346 DNLAL-----SOYDLKKYADMTDEFSSR-----LMSYATLASE 1380
 Qy 1466 TFYEGRADWCAALCYEILKCNCSKLSIRTEASQLLYFLMNNFDYTGKKSFWRTHLQVI 1525
 Db 1381 EFKSNKEMIFEL-----ESQLKLNKLESDL-----NNVE----- 1411
 Qy 1536 ISVSQIADVNGIGTRFQOSLSIINCANSRDLIKHNSFSDSVDYDLTKRITVLMATQ 1585
 Db 1412 -----KDVIRLKEESY-----HNVSJHLKLEEDPF-----KDLKRGELKYSLDN 1453
 Qy 1586 MKEHNDPEMLVDLOYSLAKSYASPEPELKTWLDMSARHIVKNGDLSAAMCYVHTLV 1645
 Db 1454 FIASIND--KIONLEVDLSKNLENTLQSF-----RLDIOKAKDKENYTLPT--- 1503
 Qy 1646 AEYLTRKAVQWEPPLPHSHSACLRSRGVFRQCTAFRVIITPNIDEASMMEDVGMQ 1705
 Db 1504 KEFSSKKCKMOSEIALMETN-----ITGVDEFVDF----- 1534
 Qy 1706 DVHNEVDVIMELLEQCADGL--WKAEVYELIA-----DIYKLIPIYEKR 1748
 Db 1535 --VNNKQSIIDSWPFLNIKDVKMOEKSYSTIEKRIINLAELGIKSPENDIFNVKIGLESFK 1593
 Qy 1749 RDPF--RLAILYDTLIRAVSKVTEVMHSGRLIGTYFRVAFPGQAQVQFTDSETPVEGPF 1807
 Db 1594 DGFELKAEIIFSNLONEAKKIQSVHLDPKNIGESINLNTL-----DEKTV 1640
 Qy 1808 EDEDEKEYIYKEPKLTPLEISQRLKLYSDRGSENVMIQDSGVNKDLSKVAYIQ 1867
 Db 1641 D-----FK-----LEKIDKVNKKTEDILIOAEVFL-----TQCKDLEK----- 1676
 Qy 1868 VTHVIFPFEKELQERKTFERSHNRPMFEMPTQIGIRQGVCEBQK----- 1917
 Db 1677 ---IFELQKLEHETTLSSMLDKVRREMVVISDSKESFEGQIEILINKNISERSEKIS 1732
 Qy 1918 --RTILTYAI--HCFPVYKRIPIVMYQHNHTDNLPIEVALIDMSKVYAEIQLCSAEVDMI 1974
 Db 1733 LYRNNIETIEENYSFSKSI-----KSDLGLE--DMLKSKLKH-----STSEIETI 1778
 Qy 1975 KLQKLQGSVSVQVNAFLADTNTKRYPDNKVILLKEVFRQFVEACGOALAVN 2034
 Db 1779 KSGLEQ-----IDKEVEFPKKNHKKELKEVDNNMLLEESKILIND 1819
 Qy 2035 ERL-----IKEDQLEYOEEMKANTREMAKELSETIMEQICPLEKTSVLPNSLHP 2085
 Db 1820 VQFNKFISEIKDNLVEYKSDLRAEFEDSYDKINFOIENQIENFPKILDSLEKNSIF 1876

RESULT 12

728676
 rhotptry protein - Plasmodium yoelii (fragment)
 CSpecies: Plasmodium yoelii
 C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #ext_change 01-Dec-2000
 C.Accession: T28676; A45521
 R.Stuba, K.A., Keen, J.K.; Ogun, S.A.; Holder, A.A.
 Mol. Biochem. Parasitol. 76, 329-332, 1996
 A>Title: Comparison of two members of a multigene family coding for high-molecular mass
 A.Reference number: Z20507; MUID:97077455; PMID:8920022
 A.Accession: T28676
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-2401 <SIN>
 A.Cross-References: EMBL:U36927; NID:g1041784; PID:g1041785; PIDN:AB41263.1
 R.Keen, J., Holder, A., Playfair, J., Lockyer, M., Lewis, A.
 Mol. Biochem. Parasitol. 42, 241-246, 1990
 A>Title: Identification of the gene for a Plasmodium yoelii rhotptry protein. Multiple c
 A.Reference number: A45521; MUID:91101660; PMID:2270106
 A.Accession: A45521
 A.Status: preliminary
 A.Molecule type: DNA

A.Residues: 2260-2401 <KE>
 A.Cross-References: GB:M34281

Query Match 2.0%; Score 214; DB 2; Length 2401;

Beet Local Similarity 18.1%; Pred. No. 0.0016; Matches 385; Conservative 336; Mismatches 760; Indels 650; Gaps 104;

Qy 54 YENVIVQKTOI-----LND---CLREMLLPYDDFQTALIRQGR-----YICST 96
 Db 597 FENILTDKTKLETKFTGLSLNNHNNKELLTYFD-----LKNLJKNENMLYKQFN 651
 Qy 97 VPAKAEER-----AQSIFVTECIKYNSDMLVNYKYEDYSSEFROLPNKVYL 145
 Db 652 EKEKAVEDIKKKNVDINKIVANIEITVTSIYN-----INDETNEIJSKSIELANTKY-L 705
 Qy 146 DKLPHVHYEVD--EVDDEDAALSGQKGIITHGMVYK-----GNNSAIS--VYMR 195
 Db 706 EKVKANVTNLNEIEKLDKDYDFQDPGEKN--IKYDENKIKNDI DTLNOKIKDSIEFTLT 763
 Qy 196 SFKRRFFHLIQLDGSYKFEFLKDLQKEPKSIFLAGLYGVSPR--NNKVRRAPELKMQD 254
 Db 764 EIKKSEHIDEIQQ-----IDKLKVPNKTMTFNEDEKEIEKLEINVEKIDKKKNIYK 818
 Qy 255 KSSYLLAADSEVEBEMWITTLNKILQNFPAAMQEKRGD--SHEDDQSKLEGGSSGL 311
 Db 819 EIDKLINISKIENDK--TSLEKLNINLSYG--KSLGNFLQOIEEKKKAETHIRYAM 873
 Qy 312 DSYLPEL--AKSAREAF-----ITLKSSESVKLYFDLPDPAQKIDFSSABEVSFPE 360
 Db 874 EAYIDDLNIRKKQOEIEKENIMNDIKMDIHKEKKNALNISHDYKIYHTTS---KNHE 929
 Qy 361 EK-----FGKRIIVKCNDSFNLQCCVANESEGP-----TNVEPFPYTL 400
 Db 930 EKISDIRKNSLKIIODFSEBSYI--NDIKKELEKLVLESQNNNDINDIYLSKIEINIVIL 987
 Qy 401 SLFDI-----YNNKISADFHVDLNFHSVROMIATSPALNNGSGPETQS 445
 Db 988 KLNKIKKIIDVKEYETDEIEKNNKINAE-----LNSBKITQQLAE--NSILKEQS 1038
 Qy 446 ALRGILHAAMQYPRQGIFFSTCCHPDLFLVARIKLVYQSLITCAIRPYMMSDSBSKAO 505
 Db 1039 KIKSTIID--NYSECIKNIT--NLKTYIYENKNI-----NTFYKAAEYNOV 1084
 Qy 506 KVLKNAQACORLQGYRMPFAMAARTLFPKASGNLDKARPSATYRQDSNKLSDNM-LK 564
 Db 1085 SLNPNINEMADTKGQYIL-----NIKKN-----NGTNNTDVNIK 1118
 Qy 565 LLADEFKPEKNAKLPVILGNLITIDNVSSDPFVYVNVSSYIPTRQFATCSKTPITFEVE 624
 Db 1119 ELKEHKKKSNNYK-----DEAGKN-----TQELIKK 1143
 Qy 625 FVPICPKHTQYTYITNHLVYYPKYLKYSOKSAKARNALICIEFKDS-----D 674
 Db 1144 NKELEFEKQEVYTLNLYVAVELKNNKEDKTKNYSB---QIKRISDAHNTFTSQADKS 1199
 Qy 675 EEDSOPLCIYGRPGGPVF--TRSAFAVLIHNOHPEVD----- 712
 Db 1200 EKKKNEIENQIRLEDVAKNNKSKALIDLQLSVEPKIRIFLKIQLARTSDDLCKETK 1259
 Qy 713 --EIKIE-----LPTQLEKHHLLTFPHV--SCDSSSKG--STKRKRVYETQV--- 755
 Db 1260 DIFETKISMLSIDQETKLIENKNILNLTLEKLSJKQKNIQDOKKELDEVNSKIKNIE 1319
 Qy 756 -----GYWMLPL-----KQGRVVTSEQHPIVSNLPSSG----- 785
 Db 1320 SNVNOHKKNYB--IGIVKINIEIAKANDQISTQKLIIPITKNLISPKANLDEGIDTNK 1378
 Qy 786 -LGVOELGMGHHYGEBEIKVNDVGKPLKISTHLSVTVYTOQHIANFPQYCOKTESGAQA 844
 Db 1379 NLGKNTNEMNNIYEFIKSYD-----LITHYLETYSKEP---ITYEQIKKKRITAO- 1426
 Qy 845 LGNELVKKYLS-----LHMEGHVMIAPFLITLNQIFRVLTATQDEAVVNV----- 891


```

Db      1427 --NELLTNKNVAKASYLDIEANEFDRIVTFHFNKLNVDVNDKFTNEYSKVNGKPDNIS 1484
Qy      892 -----TRVTHVVAQCHEEGLESLSRYVAYKAE----- 922
Db      1485 NSINNVKSTDENLNLINLQTKEMANTVSKKYYSKYEAETFINIPLANSINIQIK 1544
Qy      923 -----PYVASEKYTVHEELTKSMITILKPSADFLTSNKLKXSWFFPDV 966
Db      1545 SSSGIDLKFNINAIPLVDSQCK-----DLTFIPSEPKSETTYTMSOSYNTLADI 1597
Qy      967 LKSNAGHLIENSKVLLRNQRPASVYHNAVETVVMMLPHITQKFRNDPEASKANSL 1026
Db      1598 LKRSOELQKKEQOALWIFENRL---LHDKVQA-----TNEKDLTSLDKKKEQEI 1645
Qy      1027 AVAIFKCTFMGDFVFKOIN--NYISCPAPGDPKTLFEYKPEFLVNCHEHYPLNDP 1084
Db      1646 LNKVK-----LLHKSNELNKLSGNSQNTDLESSEKDKIKESNNYEKENIG 1696
Qy      1085 MPFGKRIQ-----RYQDLQD-YSLTDFPCRNHPLVG-LLREVGTAQ 1127
Db      1697 INPDVAKMEQFNNDIKDIKLENNYKSEKNVNSE--NNNLSQSKKLKELTNAN 1754
Qy      1128 EREVEVLIASV--KULLIKHSDPDYASRSHQARIATYLPFLGILLIENQRIINRVS 1186
Db      1755 --AEIKKIEDKIEKKGINKLLETR-----KDCMLFYKTLVETL-KIKTTDYT 1801
Qy      1187 PFPVNAGMTVYKDESLAPVNPVLPVTPQKSTLDNLSLHKLGAISGASPYTSTNINIS 1246
Db      1802 KFTTSATKSKELKIDATIS-----NSLNDP--INLQTKY-----DLNQ 1840
Qy      1247 VKNADSRGLISTDSGNSLPERNSEKSLDKHQOSSTLGNSVRCDCLOSEIKSLMNC 1306
Db      1841 INGHVSWVADATNDNNNLIEKEKATKINNLTELFTIDSKNIDAGLHNNKIQII-- 1897
Qy      1307 FLYILKMSDDDLFTWKAASSTSELMDFPTISVCLHQQYNGKRYATGMMHA-RLOQ 1365
Db      1898 -----YFN-----SEL-----HKSIDSIKOLYK-----MHAFLKLN 1924
Qy      1366 LGSLDNLSLTFNHSYGHSDADVHQSLEANIATEVCLTALDPLSLTALFKQGLADHGH 1425
Db      1925 IGH1-NKRYFDISKEDNIILOESLITANL-----NDL----- 1957
Qy      1426 NPLMKRVFVYLLCFLQKQOSETLKNVFTALSILYKPSFTEGRADCAALCYEILKC 1485
Db      1958 KEIGKISDKKKQFLHA-LSEPTIPN-FMTLKEIYDIYK-YKQID-----ELENI 2006
Qy      1486 CWSKSSITREASQILYFLMRN-----NF--DYTGKGSVTRTHLQ-----VIISQSLA 1533
Db      2007 TNEENENI--TLVYMDITLKMVKQVSLINFTVYENDSNIIKOHIDNNENNDVSKIESIE 2065
Qy      1534 DVVGIGTFRFOQSLIINNCA-----NSDLIKHTSPSSDVKDLTKRIRVYMAATQM 1586
Db      2066 TTI-----OSFOKLKLNELKAKOFOYNNINNVIT--STISQDVNDVKHISDGLTI-- 2115
Qy      1587 KEHENDPEMLVDLYQSLAKSVASTPELRKTWDSMAR-IH-----VKQGDSEAM 1636
Db      2116 ---BNE---LIGQKLEDDIKKSTVDIRGEQITKYVNPJHDVYEQTKIQQNNPKMD--- 2166
Qy      1637 CYVHTALVABYIT-RKEAVQNEPPLPISHSHACLRSGVGFROGCTAFRVYTP--NI 1692
Db      2167 ---EIDDLLOETVNVKSESELKPLTIINKDN-----VTPITISRI 2203
Qy      1693 DEBASWMEVDVGMQDVHFNEDV---LMELLEQCDGLMK--ABRYELIADI---YLLIIP 1743
Db      2204 DXYINLIK---SEYNNNDVSVYNAKKLEEDANNIIRLODTSNNLNLIDQKNEKITID 2259
Qy      1744 IYKRRDPERLAHLVDTLRAVSKVTEVHNSGRLLIGTFRVAFFGOAQQOFTDSEYD 1803
Db      2260 LKKKQGEIENRNML-QTINRE-QEIQGTETH---VANTYHH-----DINDI 2299
Qy      1804 EGFFEDEGKEYIYKPEKLTPLSEISQRLKLKYSKFGSENVKMIQDSKVPKDLDSKY 1863
Db      2300 NDVIDIND-----INDTNDINQNHQNSSDK-----KDSKRTANTGNTIRY 2340

```

Qy 1864 AYIQVTHVIFPEDEKELOERK-----TEPERS 1890
Db 2341 AGAIFGLVTFVIRIRIEKDKDEMFDKS 2371

RESULT 13

A27605
dystrophin, muscle - human
N.Alternate names: Duchenne muscular dystrophy protein
C.Species: Homo sapiens (man)
C.Date: 19-Nov-1988 #sequence_revision 27-Jun-1994 #text_change 16-Jun-2000
C.Accession: A27605; S07710; A27162; S05291; A40134; S06051; S10346; S02243; S02242; S02
R.Koenig, M.; Monaco, A.P.; Kunkel, L.M.
Cell 53, 219-228, 1988
A.Title: The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein.
A.Reference number: A27605; MUID:86194521; PMID:3282674
A.Accession: A27605
A.Molecule type: mRNA
A.Residues: 1-3685 <KOE>
A.Cross-references: GB:M18533; NID:G181856; PIDN:AA53189.1; PID:G181857
R.Rosenthal, A.; Speer, A.; Billwicz, H.; Cross, G.S.; Forrest, S.M.; Davies, K.E.
Nucleic Acids Res. 17, 5391, 1989
A.Title: Two human cDNA molecules coding for the Duchenne muscular dystrophy (DMD) locus
A.Reference number: S07710; MUID:89345106; PMID:2668885
A.Accession: S07710
A>Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: mRNA
A.Residues: 1-132, 'P', 134-622, 'I', 624-783, 'G', 785-1196, 'P', 1198-1376, 'N', 1378-1468, 'Q', 1
A.Cross-references: EMBL:X14298; NID:G30845; PIDN:CAA32479.1; PID:G30846
A.Note: This sequence was submitted to the EMBL Data Library, February 1989
R.Koenig, M.; Hoffman, E.P.; Bertelson, C.J.; Monaco, A.P.; Feener, C.; Kunkel, L.M.
Cell 50, 509-517, 1987
A.Title: Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary
A.Reference number: A50897; MUID:87273512; PMID:3607877
A.Accession: A27162
A.Molecule type: mRNA
A.Residues: 1-497 <KO2>
A.Cross-references: GB:M18533
R.Cross, G.S.; Speer, A.; Rosenthal, A.; Forrest, S.M.; Smith, T.J.; Edwards, Y.; Flint,
EMBO J. 6, 3277-3283, 1987
A.Title: Deletions of fetal and adult muscle cDNA in Duchenne and Becker muscular dystro
A.Reference number: S01263; MUID:8611512; PMID:3428261
A.Accession: S05291
A.Molecule type: mRNA
A.Residues: 404-556, 'T', 558-610, 'K', 612-622, 'I', 624-664, 'W', 665-783, 'G', 785-1137, 'P', 'N' <C
A.Cross-references: EMBL:X06178
A.Note: 475-11e and 529-Glu were also found
R.Hoffman, E.P.; Monaco, A.P.; Feener, C.C.; Kunkel, L.M.
Science 238, 347-350, 1987
A.Title: Conservation of the Duchenne muscular dystrophy gene in mice and humans.
A.Reference number: A40134; MUID:86018015; PMID:3659917
A.Accession: A40134
A.Molecule type: mRNA
A.Residues: 491-1207 <HOF>
A.Cross-references: GB:M18533
R.Blonden, L.A.J.; den Dunnen, J.T.; van Paassen, H.M.B.; Wapenaar, M.C.; Grootscholten,
Nucleic Acids Res. 17, 5611-5621, 1989
A.Title: High resolution deletion breakpoint mapping in the DMD gene by whole cosmid hyb
A.Reference number: S06051; MUID:89345155; PMID:2569720
A.Accession: S06051
A.Status: translation not shown
A.Molecule type: DNA
A.Residues: 2147-2204 <BLO>
A.Cross-references: EMBL:X15495; NID:G30829; PIDN:CAA33518.1; PID:G1335049
R.Speer, A.; Billwicz, H.; Hutch, A.; Couelle, C.; England, S.; Love, D.; Davies, K.E.
submitted to the EMBL Data Library, February 1990
A.Reference number: S10346
A.Accession: S10346
A.Molecule type: DNA
A.Residues: 2438-2480 <SPE>
A.Cross-references: EMBL:X51934
R.Chamberlain, J.S.; Gibbs, R.A.; Ranier, J.E.; Nguyen, P.N.; Caskey, C.T.

Nucleic Acids Res. 16, 1141-1156, 1988

A:Title: Detection screening of the Duchenne muscular dystrophy locus via multiplex DNA

A:Reference number: 802109; MUID:89083552; PMID:3205741

A:Status: preliminary; translation not shown

A:Accession: 502243

A:Residues: 665-722 <CHA>

A:Cross-references: EMBL:X13045; NID:g30825; PIDN:CAA31451.1; PID:g1335048

A:Accession: 502242

A:Status: preliminary; translation not shown

A:Accession: 502241

A:Residues: 2098-2146 <CH2>

A:Cross-references: EMBL:X13046; NID:g30827; PIDN:CAA31452.1; PID:g809549

A:Accession: 502244

A:Status: preliminary; translation not shown

A:Accession: 502243

A:Residues: 2147-2204 <CH3>

A:Cross-references: EMBL:X13048; NID:g30833; PIDN:CAA31454.1; PID:g1335051

A:Accession: 502109

A:Status: preliminary; translation not shown

A:Accession: 502108

A:Residues: 2305-2365, 'K' <CH4>

A:Cross-references: EMBL:X13047; NID:g30831; PIDN:CAA31453.1; PID:g1335050

R:Gilljaer, I. H. B.; van Paassen, M. H. M. B.; den Dunnen, J. J. T.; van Ommen, G. G. J. B.

A:Title: Analysis of a dystrophin gene deletion by amplification of mRNA isolated from D.

A:Reference number: 154166; MUID:91365360; PMID:1808505

A:Status: preliminary; translation not shown

A:Accession: 509071

A:Residues: 2980-2995, 'K', 2997-3028 <GIN>

A:Cross-references: EMBL:Z11860

R:Heilig, R.; Lemaire, C.; Mandel, J. L.

Nucleic Acids Res. 15, 9129-9142, 1987

A:Title: A 230bp cosmid walk in the Duchenne muscular dystrophy gene: detection of a com

A:Reference number: 809071; MUID:88067745; PMID:2825128

A:Accession: 509071

A:Molecule type: DNA

A:Residues: 'SGSHSWTHCSLIYRLPTLI', 218-277 <HEI>

A:Cross-references: EMBL:X06293; EMBL:Y00494

A:Note: sequence N-terminal of residue 218 correspond to a putative exon

R:Roberts, R. G.; Coffey, A. J.; Bobrow, M.; Bentley, D. R.

Genomics 16, 536-538, 1993

A:Title: Exon structure of the human dystrophin gene.

A:Reference number: 154186; MUID:93300536; PMID:8314593

A:Accession: 154186

A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL

A:Molecule type: DNA

A:Residues: 984-1411 <RE3>

A:Cross-references: GB:L05642; NID:g181892; PIDN:AAA74506.1; PID:g950344

A:Accession: 168509

A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL

A:Molecule type: DNA

A:Residues: 1776-1913 <RE2>

A:Cross-references: GB:L05646; NID:g181896; PIDN:AAA74507.1; PID:g950345

A:Accession: 168510

A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL

A:Molecule type: DNA

A:Residues: 2850-2979 <ROB>

A:Cross-references: GB:L05649; NID:g181899; PIDN:AAA74508.1; PID:g950346

R:Roberts, R. G.; Coffey, A. J.; Bobrow, M.; Bentley, D. R.

Genomics 13, 942-950, 1992

A:Title: Determination of the exon structure of the distal portion of the dystrophin gene

A:Reference number: 154175; MUID:92372062; PMID:1505985

A:Accession: 154175

A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL

A:Molecule type: DNA

A:Residues: 2980-3685 <RES>

A:Cross-references: GB:M66903; NID:g181881; PIDN:AAA3779.1; PID:g457519

R:Thompson, J.; Hillier, M.; Junkes, B.; Pfordt, M.; Schwinger, E.; Vosberg, H. P.

Genomics 10, 551-557, 1991

A:Title: Analysis of a dystrophin gene deletion by amplification of mRNA isolated from D.

A:Reference number: 154166; MUID:91365360; PMID:1808505

A:Accession: 154166
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 2250-2254 <RB4>
A:Cross-references: GB:S54699; NID:g235303; PID:AA19754.1; PID:g235304
R:Feener, C.A.; Koenig, M.; Kunkel, L.M.
Nature 338, 509-511, 1989
Article: Alternative splicing of human dystrophin mRNA generates isoforms at the carboxy
A:Reference number: S03902; MUID:89181947; PMID:2648158
A:Accession: S03902
A:Molecule type: mRNA
A:Residues: 'MED',12-32;3377-3408 <FEB>
C:Comment: Dystrophin is proposed to play a role in anchoring the cytoskeleton to the pl
C:Comment: Defects in dystrophin are responsible for the Duchenne/Becker muscular dystro
C:Genetics:
A:Gene: GDB:DMO
A:Cross-references: GDB:119850; OMIM:310200
A:Map position: Xp21.2-Xp21.2
A:Introns: 11/1, 31/3, 62/3, 88/3, 119/3, 177/2, 217/1, 277/3, 320/3, 383/3, 444/2, 494/4
3, 3055/1, 3075/2, 3096/1, 3121/1, 3188/2, 3217/1, 3269/3, 3325/2, 3362/3, 3408/2, 3421/1
A:Note: the list of introns is incomplete
C:Superfamily: dystrophin; alpha-actinin actin-binding domain homology; spectrin/dystrop
C:Keywords: actin binding; alternative splicing; calmodulin binding; cytoskeleton; leucit
elix
F:14-233/Domain: alpha-actinin actin-binding domain homology <ACT>
F:253-327/Region: hinge
F:338-447/Domain: spectrin/dystrophin repeat homology <SP01>
F:448-556/Domain: spectrin/dystrophin repeat homology <SP02>
F:558-667/Domain: spectrin/dystrophin repeat homology <SP03>
F:668-777/Region: hinge
F:778-828/Domain: spectrin/dystrophin repeat homology <SP04>
F:838-934/Domain: spectrin/dystrophin repeat homology <SP05>
F:938-1045/Domain: spectrin/dystrophin repeat homology <SP06>
F:1047-1154/Domain: spectrin/dystrophin repeat homology <SP07>
F:1156-1263/Domain: spectrin/dystrophin repeat homology <SP08>
F:1265-1367/Domain: spectrin/dystrophin repeat homology <SP09>
F:1372-1477/Domain: spectrin/dystrophin repeat homology <SP10>
F:1478-1568/Domain: spectrin/dystrophin repeat homology #status atypical <SP11>
F:1570-1676/Domain: spectrin/dystrophin repeat homology <SP12>
F:1678-1782/Domain: spectrin/dystrophin repeat homology <SP13>
F:1784-1875/Domain: spectrin/dystrophin repeat homology #status atypical <SP14>
F:1876-1982/Domain: spectrin/dystrophin repeat homology <SP15>
F:1984-2101/Domain: spectrin/dystrophin repeat homology <SP16>
F:2103-2208/Domain: spectrin/dystrophin repeat homology <SP17>
F:2210-2316/Domain: spectrin/dystrophin repeat homology <SP18>
F:2317-2423/Domain: spectrin/dystrophin repeat homology <SP19>
F:2424-2470/Region: hinge
F:2471-2577/Domain: spectrin/dystrophin repeat homology <SP20>
F:2579-2688/Domain: spectrin/dystrophin repeat homology <SP21>
F:2688-2802/Domain: spectrin/dystrophin repeat homology <SP22>
F:2804-2931/Domain: spectrin/dystrophin repeat homology <SP23>
F:2933-3040/Domain: spectrin/dystrophin repeat homology <SP24>
F:3041-3112/Region: hinge
F:3055-3092/Domain: WW repeat homology <WW1>
F:3080-3360/Region: cysteine-rich
F:3506-3527/Region: leucine zipper motif
F:3572-3593/Region: leucine zipper motif

Query Match 2.04; Score 213.5; DB 1; Length 3685;
Best Local Similarity 17.14; Pred. No. 0.0034;
Matches 404; Conservative 347; Mismatches 830; Indels 777; Gaps 102;

12 ETRKFTALSKPGPAALRQSEVVRGVSLLAKPKRLBPLDYENVIVOKKTOILNDCR 71
Db EDKSFSGSLNBSBVNDLRYVTALEBVLSTWLSADTL-----QAGKSTSNV-- 372
72 EML---LFPYDDFOTALLRQGRY-----ICSTVPAKAEKASQLVTECICKTYS 119
Db EVVADQCHTHBGMVMDLTAHQGRVGNLTQGSKLITGKGSLSDBEFT---VQPMNLNS 429
120 DMLLVNT---KYEDVSGEPRQLPNKKVYKDKLPVHYVEVDEVDVDED---AASIGSQK 172

Db 430 RWECLRVASMEKOSNLHRYVLMQNO--KLKELNDMLTTEBTRKREBEEPLGPDLEUDK 487
 Qy 173 GGTTHGMLYKGMNSAISVTRMSFRRRFPHLIQLDGDGYKBEFLKDLOKEPKGSIFLGF 232
 Db 488 ROVOQKRVLOEDLEQOVRNSILT-----HMVVVBESSGDHATPALEBOLK----- 534
 Qy 233 LYGVSRNNKVRFAFELMKQDKSYLLAADSEVEEMWITILINKL-----QANPE 284
 Db 535 VLGDRAAN--ICMT-----EDRWVLLODILMKQRLTEBCLFS 572
 Qy 285 AAMQERANDSHEDDEOSTLESGG--SGLDYLPBLAKSA--REAEIKLKEBSRYKLFYLDP 342
 Db 573 AMLSEK-----EDAVNKIHTTGFDONEMLSLOKLAVLKADLEKKQOSMKYSLXQ 625
 Qy 343 D-AQXLDPSASREVSFEKEGKRLVVC--NDLSNLOCCVAENEGEPTTINVEPFVYL 400
 Db 626 DLSTLTKNSVTOKTEAMLDNF-----ARCMDLVOKLEKSTRQISOAVTT--TQPSLTOT 679
 Qy 401 SLFD-----IKY-----NRKISADPHV-----DLNHSVROMIATTS 432
 Db 680 TWMEYTVTTTRQOILVKAQBELPRPPQKKQOIVDSEIRKRLVDITELHSWI----- 735
 Qy 433 PALMNGSPETOSALRGILHEAMQYPKQGISVTCRPHDI--FLVARIEKVLQGSITTHCA 491
 Db 736 -----TRG-----EAVLOSPEFAIFRKEGNFSDLEKKNVAIER-----EKA 771
 Qy 492 BPRMKSDDSKVAQOKYLK-----NAKQACORLQGYRMPFPAARTLFFDASGND 541
 Db 772 EKFRKLODSRSQALVEQVNEGVNADSTKQASELNSRWIEFC-----QLSERLNMIE 827
 Qy 542 KNARFSAIYRQ-----DSNKLSDNDMLKLLADFRKREKMAKLPLYLGNLDITIDNVS--SD 595
 Db 828 YQNNIIAFYNOLOQLEQMTTIAENWLKI-----QPTPSEPAIKSQLKICDEVNRILSG 882
 Qy 596 FRPYVNSSTIPTKQFETCSKPTTEFEVEEVPFCIRPHQTPYITTYNMLVYPRYKLYDSQ 655
 Db 883 LQPIERLKIQSIALKEGQGFPLDAD-----FVAFTTHF-----KOVFSDVQ 926
 Qy 656 KSPAKARNIAICIEFDSDEBDSQPLKCIYGRGCVFTRSAFAALVHHQNDPEFDEIX 715
 Db 927 ABEKELOTF-----DTLP-----PMRYOETMSAIRTVQO-----SEYK 961
 Qy 716 IELPLOTHEKHMLLTFEPVSCDSSKSGSTKGRDVEVTOGVSWLPLKDGKRVTSQHI 775
 Db 962 LSLP-----QLSVTDYELMEORLDELALQSLQSGLYYL-----STYK 1004
 Qy 776 PVSANLPSGYLQYQELGMRHNGPELKWDOGKPLKISTHLYSTVYTTODOLHNFQYQ 835
 Db 1005 EMSKKA PS-----EISRKYQSEFEIEG--RMWKLSSOLV-----EHC 1040
 Qy 836 QCTESGAQALGNELVYKLSLHAMEGHVMIAPLPTLNLQPLFRLTRATQEBVAVNTRYI 895
 Db 1041 QKLE-----EQMKKLKTIQNHIO-----TLKMMAEVDVFLKEBMPALGDSEIL 1084
 Qy 896 IHVVAOCH-----EBGLSHLSUYKYAYKAPRYVASEYKT----- 931
 Db 1085 KQOLQKGRLLVSDIQTIPSLSVNGGQKIKNEAPREFASRIETELKELNTQMDHMQO 1144
 Qy 932 -----VHELTKSMTTLKPSADFLTSNKLKTSWFFDVL 967
 Db 1145 VYARKKALQGLEKTVLSQKDLSEMHBMWTOAEYELERDFEYKPRDELQK-----AVEEM 1200
 Qy 968 IKSMOHLIENSKVLLRLNQRFPASVYHNAVETVNMMLPHITQKFRDNEA--SKAN----- 1023
 Db 1201 KKAKEBAQOKEAKVLLT-----ESVNSVIAQAPVQOALKEKLETLITTYQWIC 1251
 Qy 1024 -----HSLAVFIKRCFTFMDR--GVFQOINNYISCFAPGDKTLFEYK 1065
 Db 1252 TRLNKCKTLBEVWACMHELSTLEKANKMLNEVEFKLTTER-----IPGABEISEV-- 1305
 Qy 1066 FEFLLRVVCHNEHTIPLNLPMPFGKRIQRYODIQLDYSLTDFECRNHPLVGLLREVGTA 1125
 Db 1306 LPSLENTMBHSEBNDP-----NOIRILAQTLTDGVMDE-----LINELETF 1347

Qy 1126 LOEFREVELIAISVLKNNLLIKHSFDDRYASRSHQARIATVLYPLFGILLIENVRINVDV 1185
 Db 1348 NSRMELEBEAVRRK-----LLEQSIQSQOETREK 1377
 Qy 1166 SPPPVNAGMTYDESLA---LPAVNPVLTPOKSGSTLDSLHKDLGALISGIASTYTSPT 1242
 Db 1378 SLHILQESLTFIDKQLAAYIVADKYDAQMPQDAQIQSDL----- 1417
 Qy 1243 NINSVRNADSRKSLISTSGNSLPRNSENKNSLDKHOQOSTLGNVYRCOKLOSEIKS 1302
 Db 1418 -----TSHBISLEB-----MKKHQOKEAARVLSQIDVAKQLQD 1453
 Qy 1303 LLMCFYLIKMSDDALFTYNNKASTSELMDFITSEVCLHQFOYMGKRYIARTGMNAR 1362
 Db 1454 VSM-----KFRLOKRNFL-----RQESKMLDVEKMLP 1486
 Qy 1363 LOOLGSLDNSILT---FNHSYGHSDADVLHQSLEANIATEVCLTALDLSLFTLAFKNQ 1419
 Db 1487 ALETSVEQEVVQSOQNHCVN-----LYKSLSE--VKSEV-----EM 1521
 Qy 1420 LADHGHNPLMKVPVLYLCFIQKQSE-----TLKXVFTALRSLYKFPSTFYEG 1470
 Db 1522 VIKTGRQ-----IVQKQTEPNKELDERVTLKLYNELGAKVTERKQ----- 1564
 Qy 1471 RADMCALCYELIKCSNKLK--SIRTEASQLLYFLMRNPFYTGKGSFV----- 1518
 Db 1565 -----QLEKCL--KLSRKRKEMNVLTWMLAADMELT--KRSNABEGMPSLDBEV 1611
 Qy 1519 -----RTHLOVITISVQSLADVVGIGTRFOQSLSTINN-----CANSDBLI 1560
 Db 1612 AMGRATQKEIREKQKYNHLSITEVGEALKTIVGKETVLJEDKSLNSNMIAVTSABEML 1671
 Qy 1561 -----KH--TSFSSDVEDLTKRI--RTVLMATQOMKEHNDPEMLVDLOQSLKSVAST 1610
 Db 1672 NLLLEYQKMETFPDQNDVHTIKWIIQADTLIDSEKKRPOQKEDVLKTKLAKELNDIRPV 1731
 Qy 1611 PELARKTWLDSMARIHVKNGDISEAMCYVHTALVAYLTREKAAQOMEPPL--LPHSHA 1668
 Db 1732 DSTR-----DOANLMAANGD-----HKRLV-----EPQISELNRHRAA 1766
 Qy 1669 CLRSRGVFROGCTAFVITPENIDE-----EASNMEDVGMQDVAFNEV-----LME 1716
 Db 1767 ISHRIKTKASIPRLKELOFQNSDIQKLEPLEAEIQGVNLKEEDFNKDNMEDNEGTYKE 1826
 Qy 1717 LLEQCADLGMABERKELADLYKLIIPYEXRROPERLAHLYDTLHAY-----SKV 1768
 Db 1827 LLQR--GDNL-----QORITD-----ERRREBEIKIKOOLLQTHGNLKDLSQRRKKA 1872
 Qy 1769 TEVMSGRLLIGTVRAVAFQOAAQOYOPTDSETDVEGFEDDEGKEYIYKEPK--LTPLSE 1827
 Db 1873 LEISHQ-----WYOYKQRAODLLKLDLIDIEKKLALPREPRDRKIKE 1914
 Qy 1828 ISORLLK---LYSDKFGSENVKMI OOSKVPN-----KOLDSKYAVIQTHTVIPF 1874
 Db 1915 IDRELQKKKEELNNAVRQAEGLSEGDGAMAAYEPQIOLSKEMRETESKFA----- 1964
 Qy 1875 FDEKELQKRTKEFESHNIIRRFMEFMPPTQGRKQGVSEBQCKRTILITAHCPRYVKR 1934
 Db 1965 -----QERRLNFAQIHVR-----BETMMVMTEDMPELISY 1995
 Qy 1935 IPVMYQHTNDLPEVALIDEMSKVAELRQCSSAEVDMIKLOLQLOG--SVSVQVNAAG-- 1991
 Db 2053 DIHSKTAALQSATPVERVYKLOBALSQLDQOMEKVNMGYKXOQGRFDRSVEKMRFRFYD 2112
 Qy 2024 VBACQALANVERBLIKEQULEYOEBMKANYEMAKELSE-----IMHEQICP 2070
 Db 2113 IKIFQWMTLEABQFLRKTQIPEMNE--HAKYKMYLKEODIGIQGRQTVVTRTNATGEELIQ 2171

QY 2071 LEKKT--SVLNSLHIFN 2086
 Db 2172 OSSKTDASILQEKLSLN 2189

RESULT 14

114156
 kinesin-related protein - African clawed frog
 C/Species: Xenopus laevis (African clawed frog)
 C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #ext_change 11-May-2000
 C/Accession: T14156
 R/Wood, K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.
 Cell 91, 357-366, 1997

A/Title: CENP-B is a plus end-directed kinetochore motor required for metaphase chromosc
 A/Reference number: Z17893; NUID:98028574; PMID:9363944
 A/Accession: T14156
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Molecule: 1-2954 <WOO>
 A/Cross-References: EMBL:AF027728; NID:92586070; PID:92586071; PIDN:AAC60300.1
 C/Genetic: XENP-B
 C/Superfamily: centromere protein E; kinesin motor domain homology

Query Match 1.9%; Score 212.5; DB 2; Length 2954;
 Best Local Similarity 18.2%; Pred. No. 0.0027;
 Matches 424; Conservative 336; Mismatches 724; Indels 851; Gaps 111;

QY 32 SVSEVVRGSLVLAKEKLIPDYENVYQKTIQLNDCLREMLLPDYDFOTALLRQGR 91
 Db 335 STAKKIVRNT-----PQVNEVLDEALMLKRYREILD-----LKKOLE 361
 QY 92 YICSTVPAPAK-----EEEAQSLFVTECIKTYSND--WHLVNYKYEDYSGEPRQ----- 137
 Db 362 NIESSSETKAQMAAEHTQLAEIKQLHKEBEDRIMHLTNIVAS--SGESQDDQVYKX 420
 QY 138 ----LPNKVRK-----LDKLPVHVY-----EVDEEVDKD---ED 164
 Db 421 RRVTAAPGKIQNSLHASGVDFDMLSRLEGNFSKAKAFSDMPSFEIDSDVCTERSDFD 480
 QY 165 AASLSQKCGITKHCMLYGNNSAISVTMRSPKRRFHLIQLGDSYFEEFLKLOKXP 224
 Db 491 ALISMDSNDC-----IDAEWNLASKVTHREKTSIHQSMIDFGQISDVQF--HDSKSN 531
 QY 225 KCSIFLGFYGVSPFNKRYRFAFELKMDKSSYLLAADSE-----VEMEEMI 272
 Db 532 Q-----LQYLPKDSGMAEGRKASFKEKITSLOOQLOSKEEKKELVQSLELKIALEBOL 587
 QY 273 TI-----LNKILQLNFEAAMQEKRN---GESHEDDEQSLTSGS---GSGI 311
 Db 588 SVKAKNLKLVNTSRHSINAVQOTVEKEVVRKEMSVLGTSGYNASNDLQDSSVDGRKL 647
 QY 312 DSYLDELAKSAREAEIKLKSESRVYKLFYDPPAQKLDPSGAEPEVYKSFEEKGKRLV-- 369
 Db 648 SSSHDECHIRKMLBOKIYD-----LEEFLENKKKSEENDKQSSQEDMESIQICE 699
 QY 370 -----KCNLDL-----FYLQCCVAENEGPFTTNVPPFVYTLSPRIKXNRKISADPHND 418
 Db 700 AIMAEKANLBEELALMRDNPNIILENET-----LKREIADLESLENO---E 745
 QY 419 LNHFSVRQMIATTSPALMNGSGPETQSALRGILHEAMQYPRQGIFFSVCPHPDIFLVAR 478
 Db 746 TNEFFELF-----KETQKE-----HEAQI----- 764
 QY 479 IERVLQGSITTHCAEPYMKSSSDSKVAQKVLKNAKQAQRLGQYRPFAMAAATLFRKDSG 538
 Db 765 IHEI--GSI-----KKLVENAEEMNQ----- 783
 QY 539 NLDKARFPAIRQDSNKLSDNDMLKLADPRKPEKMAKLPVILGNLDTITDNVSSDPFN 598
 Db 784 NLEBDL-----ETTKLLKEOEIO-LAELEK--RADNLOKKVRFFDLSVMGDSB--- 830
 QY 599 YNNSYIPPKQFETCSKTEPTEVEBEFVDCIPGHQPYTIYTNHLVYVRYKLYKDSQK-- 656

Db 831 -----KICEB---IFOLKQSLSDAEAYTR-----DAQKEC 857
 QY 657 SPKARNIAICIEFQSDDEEDSQPLKCIYGRPGGVFRSAFAVLHHNHQNPFPDEIKI 716
 Db 858 SFLRSEN---LELKKEKEDTS-----NMTNQ--K 881
 QY 717 ELPTQLHEKHHLLTFPFVNSCDNSKSTKRDRVETQGVSWLP.LXDRGVVTSSEOHIP 776
 Db 882 EKASLSEFKQL-----ETEKSYKKQGBA-----D 905
 QY 777 VSANLPSYLGVOELGMRHYGPRIKWDGKPLI--KISHLVSTYTTTQOHLNFPQY 834
 Db 906 LQKELQSAF-----NEINYNG--LLAGVPPDLLSRV----- 936
 QY 835 CQKTESGQALGNELVYKLSLHMEGHVM-----IAPFPTLNGQFRVLTAQCE----- 885
 Db 937 --ELEKVSSEPSKOLEKLEKNALNENVTCLSEYKFLPNEVECKLQKQISASEIMULK 994
 QY 886 ---EVAVANVTVIIHVVAQCHEEGLE-----SHLSRYKVAVYKAPYVASEYKTVHEBLT 937
 Db 995 QEGHSASITIKQELIMQOSEQILQLDVETHQSKYQOT--BEQYL--EMKGNHDLF 1050
 QY 938 KSMITILKPSADFLTSTNKLKYSWFFDVLIKMAQHLIE--NSKVKLLRNQRPASVYH 995
 Db 1051 EKYIRNKSEADLLREMNELKGTWESVEKIADTYGHELEETIRQKELHKKY---PFQ 1107
 QY 996 AVETVV-----NMIMP--HIQKRPDNPASKNANHSIAVITKICFTMDGFPVKQINN 1048
 Db 1108 AMQITFPIPTPLSDSLPPSKLVEGNSQDPIEINDYNNLIALATER-----NN 1153
 QY 1049 YISCFAPDPTLTFEYKEFEFLRVVNCNHHYIPLNLP-----PFGKRIOR--YODL--- 1098
 Db 1154 IMWC-----LETENSLK-----EQVIDLNTQLOSLQAQSIKEDLDKQKPDLEEG 1199
 QY 1099 QUDSLTDFCRNHFPL-----VGLLREYGTALQEFREYRLAISVLKNLIKH--- 1147
 Db 1200 EVKLLLEMEELKGLTDSQLSIEKIQLENTLEVTETKQLTQCE-----EMKQITIERMEL 1252
 QY 1148 --SPDRYASRSHQARITLVLP.LGILLIENVQRIINRVADVPPFNAGMTYKDSLALPA 1205
 Db 1253 QTFEEDLAEHD-----SLKQDISENIEQ-----SLETQD--IRA 1286
 QY 1206 VNPLVTPQKSGTLDNSLHKLDLGAISGASYTTSTP-----INSVNRADSG--- 1254
 Db 1287 AOEBLEBEK--QLVDSFQQLDCSVGISSEPHDAVANOEKVSLGEVNSLSQSEMLRGIRD 1344
 QY 1255 -----SLIS-----TDGNSLPERNSEKNSLDKHOQOSTLGNSVVRCDKLDQ 1297
 Db 1345 ELQTSKALVSELELLRAHVKSVEGENL--EITKLNGLER-----BILGKS-----EE 1391
 QY 1298 SEIKSLMCFIYLKSMDDALPTYMANKASISLMDPFTTISEVCLHQCQYNGKRYIATG 1357
 Db 1392 SE-----VKSMLDEN-----KEDNNKLUKSE----- 1411
 QY 1358 MHAARLQOLGSLDNLSTFNHSYGS-----DADVLHQSLEANLATEVC-----LTALD 1406
 Db 1412 -----QAEYSKQENQFSLAEVFSQKLVDEIYVAKQKAAEBLEIKDDYFLVQTA 1467
 QY 1407 TLSLETLAFKQNLADHGHPMLKVPDVYL.CFLQKQSEFALKNVFTALBSLYKFPST 1466
 Db 1468 NTNLVGEKLETPLOQDHEDSIDRRSEMEI---KVLGKLENNQVLLER----- 1514
 QY 1467 FYEGADKCALCYEILK-----CNSKLSISRTASQLYFLMKNPNPYTGK- 1514
 Db 1515 LOEBEKLGSNKL--EILQKEMETSVLTKDLQOKLESLSB-----NIIKEMDITTLKH 1567
 QY 1515 KSFVRTHQVILISVSQILADVVGIGTRFOQSLIINN--CANSDR--LIKHTSPSSD-V 1569
 Db 1568 HSDTQAOQ-----KTQEBQLAQKLAIAADNCPITQEKESALCV 1609
 QY 1570 KDLTKRITVLMATQAQMKHENDPEMLV---DLQYSLAKSYASTPELAKTMDLSMA--- 1622

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:02:46 ; Search time 21.9153 Seconds

(without alignments)
3987.653 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936

Sequence: 1 MSQPLPLPASAEKTRKTRAL.....ISGTPSTWVHGMTSSSSSV 2107

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	213.5	2.0	3685	1 DMD_HUMAN	P11332 homo sapien
2	203	1.9	2077	1 TEGU_HSV6G	P30002 human herpes
3	199	1.8	3210	1 CENF_HUMAN	P49454 homo sapien
4	196.5	1.8	2748	1 NDM1_YEAST	Q00402 saccharomyc
5	195	1.8	2077	1 TEGU_HSV6U	P52340 human herpes
6	193	1.8	1325	1 G160_MOUSE	P55337 mus musculu
7	190.5	1.7	4568	1 DYHC_CAEEL	Q19020 caenorhabdi
8	187.5	1.7	2869	1 RBP1_PLAVB	Q00798 plasmodium
9	186	1.7	1818	1 HMM2_MYCPN	P75471 mycoplasma
10	185	1.7	2663	1 CENB_HUMAN	Q02324 homo sapien
11	185	1.7	1557	1 DVAL1_DICV1	Q24702 dictyocaulu
12	184	1.7	4385	1 YP73_CAEEL	Q09222 caenorhabdi
13	183.5	1.7	1901	1 AK11_HUMAN	Q90484 homo sapien
14	183	1.7	5327	1 ACF7_MOUSE	Q94920 mus musculu
15	181.5	1.7	1957	1 YD66_SCHPO	Q10411 schizosacch
16	181	1.7	2014	1 YUJ7_YEAST	P39526 saccharomyc
17	181	1.7	3680	1 DMD_CANFA	O97592 canis famli
18	179.5	1.6	1966	1 MYSE_CAEEL	P02566 caenorhabdi
19	178.5	1.6	1679	1 Y109_YEAST	P40457 saccharomyc
20	176.5	1.6	2273	1 HPA1_YEAST	P32874 saccharomyc
21	175	1.6	1805	1 HMM2_MYCGE	P47460 mycoplasma
22	172	1.6	3056	1 ATM_HUMAN	Q13315 homo sapien
23	171.5	1.6	3678	1 DMD_MOUSE	P11531 mus musculu
24	171	1.6	1151	1 XPO4_MOUSE	Q99310 mus musculu
25	169	1.5	1526	1 MYB2_SCHPO	Q90416 schizosacch
26	168	1.5	3660	1 DMD_CHICK	P11533 gallus galli
27	167.5	1.5	1875	1 MLPI_YEAST	Q02455 saccharomyc
28	167	1.5	1151	1 XPO4_HUMAN	O99302 homo sapien
29	166	1.5	2136	1 YCF2_MARPO	P09975 marchantia
30	165.5	1.5	2704	1 BRP1_HUMAN	Q03801 homo sapien
31	165	1.5	2469	1 TEGU_HSV5A	Q01056 herpesvirus
32	164	1.5	2059	1 TEGU_HSV7J	P52362 human herpes
33	164	1.5	3911	1 AKA9_HUMAN	Q99996 h-a-kinase

34	164	1.5	4092	1 DYHC_YEAST	P36022 saccharomyc
35	164	1.5	4128	1 PRKD_MOUSE	P97313 mus musculu
36	162.5	1.5	2875	1 RRP1_TSWV1	P23976 tomato spot
37	161.5	1.5	5430	1 ACF7_HUMAN	Q94920 mus musculu
38	160	1.5	1928	1 MYB1_YEAST	P03964 saccharomyc
39	159	1.5	2026	1 CYA1_YEAST	P03678 saccharomyc
40	159	1.5	2472	1 SPCN_RAT	P16086 rattus norv
41	158	1.4	4563	1 APB_HUMAN	P04114 homo sapien
42	157	1.4	1420	1 YMB8_YEAST	Q03496 saccharomyc
43	157	1.4	2144	1 BP28_HUMAN	Q9H583 homo sapien
44	157	1.4	2349	1 TPR_HUMAN	P12270 homo sapien
45	157	1.4	3080	1 POLG_ZMYVC	P18479 z genome po

ALIGNMENTS

RESULT 1	DMD_HUMAN	STANDARD;	PRT; 3685 AA.
ID	DMD_HUMAN	P11532; Q14169; Q14170;	
AC	P11532; Q14169; Q14170;		
DT	01-OCT-1989 (Rel. 12, Created)		
DT	01-OCT-1989 (Rel. 12, Last sequence update)		
DT	15-0UN-2002 (Rel. 41, Last annotation update)		
DE	Dystrophin.		
GN	DMD.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OK	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=88194521; PubMed=3282674;		
RT	Koenig M., Monaco A.P., Kunkel L.M.;		
RT	"The complete sequence of dystrophin predicts a rod-shaped		
RT	cytoskeletal protein.";		
RT	Cell 53:219-228(1988).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=89345106; PubMed=2668885;		
RT	Rosenthal A., Speer A., Billowitz H., Cross G.S., Forrest S.N.,		
RT	Davies K.E.;		
RT	"Two human cDNA molecules coding for the Duchenne muscular dystrophy		
RT	(DMD) locus are highly homologous.";		
RT	Nucleic Acids Res. 17:5391-5391(1989).		
RN	[3]		
RP	SEQUENCE OF 1-497 FROM N.A.		
RX	MEDLINE=87273512; PubMed=3607877;		
RT	Koenig M., Hoffman B.P., Bertelson C.J., Monaco A.P., Feener C.,		
RT	Kunkel L.M.;		
RT	"Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and		
RT	preliminary genomic organization of the DMD gene in normal and		
RT	affected individuals.";		
RT	Cell 50:509-517(1987).		
RN	[4]		
RP	SEQUENCE OF 404-1137 FROM N.A.		
RX	MEDLINE=88111512; PubMed=3428261;		
RT	Cross G.S., Speer A., Rosenthal A., Forrest S.M., Smith T.J.,		
RT	Edwards V., Flint T., Hill D., Davies K.E.;		
RT	"Deletions of fetal and adult muscle cDNA in Duchenne and Becker		
RT	muscular dystrophy patients.";		
RT	EMBO J. 6:3277-3283(1987).		
RN	[5]		
RP	SEQUENCE OF 665-722; 2098-2204 AND 2305-2366 FROM N.A.		
RX	MEDLINE=89083552; PubMed=3205741;		
RT	Chamberlain J.S., Gibbs R.A., Ranier J.A., Nguyen P.N.,		
RT	Caskey C.T.;		
RT	"Deletion screening of the Duchenne muscular dystrophy locus via		
RT	multiple DNA amplification.";		
RT	Nucleic Acids Res. 16:11141-11156(1988).		
RN	[6]		
RP	SEQUENCE OF 2147-2204 FROM N.A.		
RX	MEDLINE=89345155; PubMed=2569720;		

RA Blonden L.A.J., den Dunnen J.T., van Paaseen H.M.B.,
 RA Wapenaar M.C., Grootecholten P.M., Ginjaar H.B., Bakker E.,
 RA Pearson P.L., van Ommen G.J.B.;
 RT "High resolution deletion breakpoint mapping in the DMD gene by whole
 RT cosmid hybridization.";
 RL Nucleic Acids Res. 17:5611-5621(1989).
 RN [7]
 RP SEQUENCE OF 2305-2364 FROM N.A.
 RA Hutch A., Will K., Speer A., Bauer D.;
 RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP REVIEW ON DMD POINT MUTATION VARIANTS.
 RX MEDLINE=95038763; PubMed=7951253;
 RA Roberts R.G., Gardner R.D., Bobrow M.;
 RT "Searching for the 1 in 2,400,000: a review of dystrophin gene point
 RT mutations";
 RL Hum. Mutat. 4:1-11(1994).
 RN [9]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=94320940; PubMed=8045556;
 RA Rimsland F., Reiss J.;
 RT "Microlesions and polymorphisms in the Duchenne/Becker muscular
 RT dystrophy gene.";
 RL Hum. Genet. 94:111-116(1994).
 RN [10]
 RP VARIANTS DMD ARG-54.
 RX MEDLINE=94004962; PubMed=8401582;
 RA Prior T.W., Papp A.C., Snyder P.J., Burghes A.H.M., Bartolo C.,
 RA Sedra M.S., Western L.M., Mendell J.R.;
 RT "A nonsense mutation in the dystrophin gene in a Duchenne muscular
 RT dystrophy patient.";
 RL Nat. Genet. 4:357-360(1993).
 RN [11]
 RP VARIANTS DMD HIS-365; TRP-2191 AND ARG-2937.
 RX MEDLINE=95352525; PubMed=7849724;
 RA Nigro V., Nigro G., Bosposito M.G., Comi L.I., Molinari A.M.,
 RA Puca G.A., Pollano L.;
 RT "Novel small mutations along the DMD/BMD gene associated with
 RT different phenotypes.";
 RL Hum. Mol. Genet. 3:1907-1908(1994).
 RN [12]
 RP ALTERNATIVE SPLICING (DYSTROPHIN-1 AND -2).
 RX TISSUE=Retina;
 RA White R.A.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [13]
 RP ALTERNATIVE SPLICING (DYSTROPHIN-3).
 RX TISSUE=Brain;
 RX MEDLINE=89181947; PubMed=2648158;
 RA Feener C.A., Koenig M., Kunkel L.M.;
 RT "Alternative splicing of human dystrophin mRNA generates isoforms at
 RT the carboxy terminus";
 RL Nature 338:509-511(1989).
 CC -1- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
 CC PLASMA MEMBRANE.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 4 ISOFORMS ARE PRODUCED BY
 CC ALTERNATIVE SPLICING; THE 3 SHORTER VARIANTS ARE CALLED
 CC DYSTROPHIN-1, -2 AND -3.
 CC -1- DISEASE: Defects in DMD are the cause of Duchenne muscular
 CC dystrophy (DMD) and Becker muscular dystrophy (BMD). DMD is the
 CC most common form of muscular dystrophy; a sex-linked recessive
 CC disorder. It typically presents in boys aged 3 to 7 year as
 CC proximal muscle weakness causing waddling gait, toe-walking,
 CC lordosis, frequent falls, and difficulty in standing up and
 CC climbing up stairs. The pelvic girdle is affected first, then the
 CC shoulder girdle. Progression is steady and most patients are
 CC confined to a wheelchair by age of 10 or 12. Flexion contractures
 CC and scoliosis ultimately occur. About 50% of patients have a lower
 CC IQ than their genetic expectations would suggest. There is no
 CC treatment. BMD resembles DMD in hereditary and clinical features
 CC but is later in onset and more benign.
 CC -1- DISEASE: Defects in DMD are a cause of X-linked dilated
 CC cardiomyopathy (XLCM).

CC -1- MISCELLANEOUS: THE DMD GENE IS THE LARGEST KNOWN GENE IN HUMANS.
 CC IT IS 2.4 MILLION BASE-PAIRS IN SIZE AND COMPRISES 79 EXONS.
 CC -1- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
 CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIBRININ,
 CC ABP-120, ABP-180, OR BETA-FODRIN).
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
 CC -1- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.
 CC -1- SIMILARITY: CONTAINS 1 WW DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 ZF-TYPE ZINC FINGER.
 CC -1- DATABASE: NAME=DMD; NOTE=Dystrophin Mutation Database;
 CC WWW="http://www.dmd.nl/database.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).
 CC -----
 CC EMBL, X06179; CAA29545.1; ALT_SEQ.
 CC EMBL, X06178; CAA29544.1; -.
 CC EMBL, X14298; CAA32479.1; ALT_SEQ.
 CC EMBL, X15495; CAA33518.1; -.
 CC EMBL, X54820; CAA38589.1; -.
 CC EMBL, X13045; CAA31451.1; -.
 CC EMBL, X13046; CAA31452.1; -.
 CC EMBL, X13047; CAA31453.1; -.
 CC EMBL, X13048; CAA31454.1; -.
 CC EMBL, U27203; AAA86115.1; -.
 CC EMBL, X15148; CAA33245.1; -.
 CC PIR, A27605; A27605.
 CC PIR, A27162; A27162.
 CC PIR, S05291; S05291.
 CC HSSP, P46939; IQAG.
 CC Genew; HGNC:2928; DMD.
 CC MIM; 300377; -.
 CC MIM; 310200; -.
 CC MIM; 300376; -.
 CC MIM; 302045; -.
 CC InterPro; IPR001589; Actbind_actnin.
 CC InterPro; IPR001715; Calponin-like.
 CC InterPro; IPR002017; Spectrin.
 CC InterPro; IPR002349; WW_Rsp5_WMP.
 CC InterPro; IPR001202; WW_Rsp5_WMP.
 CC InterPro; IPR000433; Znf_ZZ.
 CC Pfam; PF00307; CH; 2.
 CC Pfam; PF00397; WW; 1.
 CC Pfam; PF00435; Spectrin; 22.
 CC Pfam; PF00569; ZZ; 1.
 CC PRINTS; PR00403; WMDOMAIN.
 CC SMART; SM00033; CH; 2.
 CC SMART; SM00150; SPEC; 21.
 CC SMART; SMO0456; WW; 1.
 CC SMART; SMO0291; Znf_ZZ; 1.
 CC PROSITE; PS00019; ACTININ_1; 1.
 CC PROSITE; PS00020; ACTININ_2; 1.
 CC PROSITE; PS50021; CH; 2.
 CC PROSITE; PS01159; WW_DOMAIN_1; 1.
 CC PROSITE; PS50020; WW_DOMAIN_2; 1.
 CC PROSITE; PS01357; ZF_ZZ_1; 1.
 CC PROSITE; PS50135; ZF_ZZ_2; 1.
 CC Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
 CC Repet; Disease mutation; Alternative splicing; Zinc-finger.
 CC DOMAIN 1 240
 CC DOMAIN 1 119
 CC DOMAIN 1 134 237 CH 1.
 CC REPEAT 134 237 SPECTRIN 1.
 CC REPEAT 448 447 SPECTRIN 2.
 CC REPEAT 559 556 SPECTRIN 3.
 CC REPEAT 719 828 SPECTRIN 4.
 CC REPEAT 830 934 SPECTRIN 5.

Query Match 2.0%; Score 213.5; DB 1; Length 3685;
 Best Local Similarity 17.1%; Pred. No. 0.0013;
 Matches 404; Conservative 347; Mismatches 830; Indels 777; Gaps 102;

12 EFRKTRALSKGTAAELRQSVSEVVRGVLAKPELLELDYENVYQKTOILLDCIR 71
 326 EKSFGSSLSMESEVNLDRYQALREVLISLAEEDTL-----QAQGEINDV-- 372
 72 EML-----LPEYDDPQTALILRRGRY-----ICSTVPAKAEBAQSLFYRECIKTYS 119
 373 EYVKQGFTHBEGYMDLTAHQGRVGNILQSGKLIGTGKL3EDETE---VOEQWNLNS 429
 120 DMHLVNY-----KYEDYSGEFRQLPNKVVLDLKYVHYVEVDEVDKED---AASIGSQK 172
 430 RRECLRVASMEKQSLHRYLMDLQNG--KLKELNMLTERTTRKMEERPGPRLBEX 487
 173 GGIYHGMVYKGNNSAISVTRKSRFRPFHLIQLDGSGYKPEFLDKQEPKSGIPLGF 232
 488 RYVQCHKVLQEDLEQEOVRVNSLT-----HMVVVDESSGDHATVAALEEQLK----- 534
 223 LYGVSFRNNKVRRAFAELKMDKSSVLLADSEVEHEWITLNLKIL-----QLNPE 284
 535 VLGRMAN--ICRMT-----EDRWVLQDILLKMQRLTERQCLFS 572
 285 AAMQEKRGDSDHEDEQSKLESG--SGLDYLPFLAKSA-REAEIKLKSESHVKLFYLDP 342
 573 AMLSEK-----EBAVKKIHTTGFQDQENMSSLOKLAIVLKADLEKKQSGKLYSLKQ 625
 343 D-AQDLRPSAEPVKSFEKFGKILVKC-NDLSFNIQCCVAENEDEGTTNVEFPYTL 400
 626 DLSTLKMKSVYQKTEAWLDNF-----ACMDNLVQKLEKSTAQISQAVTT--TOPSLQOT 679
 401 SLFD-----IKY-----NRKLSADPHV---DLNHSFRQMIATTS 432
 660 TMEVTVTTTREQILLVKAQELPPPPQKRRQLTVSEIKRLLDVTTELHSMI---- 735
 433 PALMNGSGPETQSALRGILHEAMQYPKQGFISVTCPEHPI--FLVAREIKVQSGITHC 491
 726 -----TRS-----EAVLOSPEFAIRFKSGNPSULKEXNALIER-----EKA 771
 492 EPMYKSSDSSKVAQKVLK-----NAKQACRLQYRMPFMAAATFLFKASGND 541
 772 EKFRKLQDASBSAQALVEQMGNEGVNADSIKQASEQNLKHWIEFC---QLSERLNMLE 827
 542 KNARSATYRQ---DSNKLSDMDLKLADFRKBEKAKLPEVILGNDITIDNVS--SD 595
 828 YQNNIIAFNQLOOLEQMTTAEMLKI-----OPTPSBPFAIKSOLKICQDEVNRLSG 882
 596 PPNVNSSYIPTKQFETCSKTPITEFEVEEFVPCIRHQTQYTIYNHLVVPKYLKYSQ 655
 883 LQPIREKLQISIALKEQKQPMPLDAD-----FVAFNNHF---KQVPSDVQ 926
 656 KSFAPARNIALCIEFKDSDDEDSOPLKCIYGRPGGVPFTSAPAAVLHNNQPEFYDEIK 715
 927 AREKELQITF-----DTLP-----PMRYQSTMSAIRTWVQO---SEIK 961
 716 IELPQOLEKHHLLTFPHVSCDSSKSGTKRDVETQVGSWPLKLDGAVVTSEQHI 775
 962 LSIPT-----QLSTYDYEIMEORGELOALQSSLOEQSGLYYL-----STYVK 1004
 776 PVSANLPSGYLGYOELGMRHYGPRIKWDGKPLIKTISTHLVSTVYVTDQHLNHPQVC 835
 1005 ESKSKAPS-----EISRKYQSEFEIEG--RKKLSSQV-----EHC 1040
 836 QKTESGAOLAGNELVYKLSLHAMEGHVIAFLPTLLNQLFRVLTTRATQOEVAVANVTRY 895
 1041 QLEF-----EQMKLRKIQWHIQ-----TLKWMMAEVDFVLKEMFALGSEIL 1084
 896 IHVVAQCH-----BEGLESHLSRYKVAKYKPEPYASEKT----- 931
 1085 KKQKQCHLLVSDIQTIQPSLNSVNEGQKIKNEABEPFASRLTELKELINTQMDHMCQ 1144

932 -----VHEELTKSMTTLKPSADPLTSLNLLKYSWFFPDL 967
 1145 VYAREKALKGGLKLEKTVSLQKDLSEHWEWMTQAESEYLERDREYKTPDLQK---AVREM 1200
 968 IKSMAGHLIENSKYKLYLRNQRPASYNHNAVETVNNMLMPLHTQKRPDPEA-SKAN--- 1023
 1201 KRAREBAQKBAKVKYLT-----ESVNSVIAQAPVQAEALKSLLETLTNNYQWLC 1251
 1024 -----HSLAVFIKRCPTFMDR-GFVFKQJNNYISCAPDQPKLPEFYK 1065
 1252 TRLNCKCTLEBWACHMELLSYLEKANKMLNEVEFKLTEN-----IPGAEISISV- 1305
 1066 FEFLRVVNCBHRYIPLNMPFPGKRIQRYODLODLSLTDEFCRNHFLVGLLEEVGTA 1125
 1306 LDSLLENLRHSEDDN-----NQRIILAQTITDGGVME-----LINEBELTF 1347
 1126 LOEFREVRLLISVLKULLIHSFDPRVASKSHQARIYLYPLFGLLIENVQRIINADV 1185
 1348 NSRMELEHEBAVRQK-----LLEOSIQSQAQETEK 1377
 1186 SPFPVNAQMTVKDSIA---LPAVNPLVTPQKSGTLDNSLHKDLGALSGIASPYTSTP 1242
 1378 SLHLIQESLTIIDKQLAAYIADKVDAAQMPQEAQIKQSDL-----LEOISQSAQETEK 1417
 1243 NINSVRNADSRGSLISTDGSNLSPERNSEKNSLIDKHOQSTLANSVVRCDKLDQSEIKS 1302
 1418 -----TSHEISLEB-----MKKNHQGBAQRVLISQIDVQKQLD 1453
 1303 LMKCFYLYLKSMSDALTFTYNNKASTSELMDFITISEVCLHQFOYMKRYIARTGMHAR 1362
 1454 VSM-----KFRLLQKPAFNPFL-----RIQESKMLIDBEVQMLP 1486
 1363 LQQLGSLDNLSTL---FNNSYGHSDADVHOSLLEANITAEVCLTALDLSLFTLAFNQOL 1419
 1487 ALETYSVQAEVVOQOLNHCN-----LYKLSB--VASEV-----EM 1521
 1420 LADHGHNLMKVPDVVLCFLQKQSE-----TALKRVFALBSLIYKFPSTPYEG 1470
 1522 VIKGRQ-----IYQKKQTEMPKELDRVATLKLHYBELGAKVTERQ----- 1564
 1471 RADMCALCYEILKCKNSKLS-STRTASQULLYPLMRNNPYTGGKFPV----- 1518
 1565 -----OLEKCL--KLSRKVRKEKENVLTWELAAIDMELT--KRSAGEWPSNLDSBV 1611
 1519 -----RTHLOVITISVQQLADVIGIGTRQOSLSIINH-----CANSDLI 1560
 1612 AMGKATQKEIKQKVLHLSITEVBALKTVGKKEVLEVDLSLNSNMIAVTSABEML 1671
 1561 -----KH-TSFSSDVKDLTKRI--RTVLATAQMKENDEPMKVLQYSLAKSVAST 1610
 1672 NLLLEYQGHMETFQONVDHITKWIIQADTLDESEKPKPQKQKEDVLRKALNDIRKPV 1731
 1611 PELRKTWLDNARIHVKNKGLISEAAMCYHVATLVAEYLTRKEAVQWPPPL--LPHSHA 1668
 1732 DSTR-----DQANIMANRGD-----HCRKLV-----BQJISELHNRPA 1766
 1669 CLRSRGVFPQGCATPAPVITPND-----EASMEDVGMQDVHFNENDV-----LME 1716
 1767 ISHRKTKGKASIPKLEQFNSDIQKLEPLEASIQOVNLEKEDFNKDNMNEGTYKE 1826
 1717 LLEQCADGLMAERYELADIYKLLIPYERKRPFERLAHLYDTLHRAV-----SKV 1768
 1827 LLQR-GDVL-----QQRITD-----ERGREIKIKOULLQTKHNAKULRSORRKA 1872
 1769 TEVMSGRLLGTYFRVAFGOAAYOFTDSETVDEGEFEDEDEGEYIYKSPK-LTPUSE 1827
 1873 LEISHQ-----WYQYKQADOLLKCLDIDIRKTLASLPEPRDERKIKE 1914
 1828 ISQRLK---LYSDKRGSEVNMKIQDSGKNP-----KOLDSYAYIQTHTVHP 1874
 1915 IDRELQKKKEELNVRQQAEGLSGDGAAMAVEPQIOLSKRMVIESKFA----- 1964
 1875 FDEKELQERKTEPERSHNIRFMFEMPEPTQYQKQGVGEQCKKRTILTAIHCPYVAKR 1934

Db 1965 -----QFRNLNFAQIHTVR-----EETMMNTMEDMPELEISY 1995
 Qy 1935 IPVMYQHNDLNIENVAIDEMSKKVAELRQCGSSAEVDIMIKLQKGS-SVSQVNAAG-- 1991
 Db 1996 VBSTYL--TEITHVSQALREV--BQLNAPDLCKDEDFLFOESLKNIKDSLQSSGRI 2052
 Qy 1992 -----PLAYARAFIL--DTNTKRYPDNKVLLKEV--FROP-- 2023
 Db 2053 DIHSKKTALQSGATFVERVKLQDALSQLDFQWKEKNKVKQKQKGFDRSVKMRFRPHD 2112
 Qy 2024 VEAACGALAVNELIKEDOLEYOEMKANTREMAKLSL-----IMHEQLCP 2070
 Db 2113 IKTFNOMLTBAEQFLKTKQIPENWME-HAKYKMYLKLQDQIGQRTVVRVTLNATGEIITQ 2171
 Qy 2071 LEEKT--SVLPNSLHIEN 2086
 Db 2172 QSKSDASILQEKGLSLN 2189

RESULT 2

TEGU HSV6G STANDARD; PRT; 2077 AA.
 AC P30002;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Large tegument protein.
 GN U31.
 OS Human herpesvirus (type 6 / strain GS) (HHV6).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Betaherpesvirinae; Roseoloviruses.
 RN NCBI_TaxID=10369;
 RX MEDLINE=91374623; PubMed=1654455;
 RA Joseph S.F., Ableseth D.V., Salahuddin S.Z., Jagodzinski L.L.,
 RA Wong-Staal F., Gallo R.C.;
 RT Identification of the human herpesvirus 6 glycoprotein H and
 RT putative large tegument protein genes.";
 RL J. Virol. 65:5597-5604 (1991).
 CC FUNCTION: TEGUMENT PROTEIN.
 CC SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
 CC -1 SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
 CC EHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL, S57540; AAB19786.1; --
 DR PIR; HA0511; WZBE24.
 SQ SEQUENCE 2077 AA; 239909 MW; 28E297FDC2FAD2AB CRC64;

Query Match 1.9%; Score 203; DB 1; Length 2077;
 Best Local Similarity 18.0%; Pred. No. 0.0022; Indels 500; Gaps 80;
 Matches 294; Conservative 268; Mismatches 567;

Qy 694 TSSAPAVLHHNQPEYDEIKIELPTQLHEKQ---HLLLTFFHVSNNSSKSTKKRD 749
 Db 60 TGTGLTALMNNSSQGY--RLPTEIPKRIHSRGVYGHLSRPFNTLD-----TQK-- 109
 Qy 750 VVETQVGVSWPLDKGAVVTSEQIIPVANSIPSGYLQOEL--GMGHYGEPEIKWDG 806
 Db 110 -IDNEV-----YLGILIDILYKSKSRN--GPTFAVITI 138
 Qy 807 GKPLKISTHLVSTVYTOPDQ---LHNFPQYCKCKTESGAOL---GNE-----LVK 851
 Db 139 G-VLSKALFPLNNTLTLFDSHPTEKREATAIYICQIEVEYELLTTHGTEGYYDASLIF 197

Qy 852 YLK-----SLHANEGHVMI-----AFPLPTLNLQFRVLTATQOEAVANVTR 893
 Db 198 FIETSNLSLSSHDAELLILKTYKDPDIALDKFSESTIEHKTKDDIGSQODLVADKTT 257
 Qy 894 VTIHV--VAQCHEBGLSHLSYVYKAVKAEYVASEKTYNHE--ELTKSMITLIKPSAD 949
 Db 258 DLEHAPHRKKNSHLELELNDKKKQDASLYVATEVDLLPSFEYLSQFOSLFHDLS 317
 Qy 950 F--LTSNKLKYSWFFPVLLIKSMAQHLIENSKVTLNORFPASVYHNAVETVMMLMPH 1007
 Db 318 FPIKMSN--FNWITIY-----LQDSPII--PNOGPATPF--LNNRVHLLCQ 357
 Qy 1008 ITQKE---RDNPASKNANSHLAVFIRCFETMDRGFVFKQ-----INN 1048
 Db 358 IIDVFVGSGTNDSSKQOQOT--IFINYLLPFKQSFVFMALAACQBNNDIIFYNN 415
 Qy 1049 YISCFAPDQPTLFYKFE-----FLAVVCH--HEHT-----IPLNMPPECKGIQ 1093
 Db 416 YL-C-----KTTYFRTLERILSKFLAIVNDHDKGHYEWKSWTQMFQCKPKLDIE 468
 Qy 1094 RYQDLQDYSLDFE-----CRN---HPLVGLLRVGTALQEFREVRLLAISVLKNLLIK 1146
 Db 469 NYLKAYVDHNPVKGAFHEFTCLNKAETKVAVALNEKKEKIEQALIREKNSFQALSN---- 524
 Qy 1147 HSFDDRYSRSHQARIATLYPLFGLLIENYQRIWVRDVSPPVNAGMTVXDESIALPAV 1206
 Db 525 --FIDKLGs-----TPALPIESENVHKHTSDITGIVRPTI--ESTIELPNI 568
 Qy 1207 NPLVTPQKSGTLDNSLHDLGALSIPPTTS--PINSVRNDSGSLSTD----- 1260
 Db 569 SLTNNTOQIS--LDKQNLNKLNTNTHTLTKKTKYVQDDYNNI---AAGFMPEVELNCLF 623
 Qy 1261 -----SGNSLPERNSEKNSLDDKQOQSTLGSVYRCQKQDSIKSLIM 1305
 Db 624 AYLVNLYNIEVLKSGSLNITVVLQEVKLYNDQPIARFGTSHNINNLN----- 675
 Qy 1306 CELYLKMSMDALFTYNNKASTSELDMPFTISEVCLHQFOYMG--KRYIARTGMWA 1361
 Db 676 -FTLSIRKQFVD---FNSQKPSD---BASEILAIESTILDPSKNTVNIEMIKS 725
 Qy 1362 RLQQLGSLDNGSLTFNHS-----YGHSDADVHLQSLLEANIATEVCLTALDTLSPTL 1413
 Db 726 QLEELGKKEISTEKKQTAITTKQILQGOELTPPIYDFLHLHSAVNLPTTIVYKNLHLHFI 785
 Qy 1414 AFKNQDLADHGNPMLKVFVDVYLCFLQKHQSEFALKNVPTALRSILYKFPSTFEGRAD 1473
 Db 786 LEQRDIAMTLHDKI--QSILDIYVDDML--NDIYPEQTFSTVLPVLFPFNSTK----- 838
 Qy 1474 MCAALCYEIL-----KCCNSK-----LSIRTEASQLLYFLMRNPNFYTGKKSFVR 1519
 Db 839 --TALFESVLTIRQLAKKCANMLKTLDEPDLAQFITNSBQLQNNMRQH--GKK----- 889
 Qy 1520 THLOVITVSQILADVIGIGTRFQOQSIIINNCCNSRLIKHSFSSDVYDUTKIRITV 1579
 Db 890 -----IPLT-----MGHIKFLYSQKII-----TTEBKQNIQIRACTV 920
 Qy 1580 L-----MATA-----OMKE--HENDEPMLVDLOYSLAK 1605
 Db 921 VITSBELTALFATPHTHALQTCRPLDKALQRMHEQOMQOTABNDKHHLLITRSTLEK 980
 Qy 1606 -----SYASTP-----ELRKTWL-----DSMARIHVKNGLSEAMCYUHVATLV 1645
 Db 981 RLNDILLILKQGFSELTETMLNLLETFLKQLODNNVNIHFTHALLPVLKQIETITSKII 1040
 Qy 1646 AEYLTRKAVQWEPPLPHSHS-----ACLRSRGGVF-----RQOST 1683
 Db 1041 SDVI---EKILIKTPLNBEQMSKEQKYTPLLSFLSKKTKTTFCTEDVKTEDIDQWQSKIT 1097
 Qy 1684 AFRVITPNIDEBASMMEDVGMQDVHFNEDVAMEL-----LEQCADGLMAERYELIAD 1736
 Db 1098 FLKIKATISANGTRLSHSIYQOELNLYBERITELRKETNNKKEQSLSKYALAEK----- 1151
 Qy 1737 IYKLIIPYEKRRDPEBLAHLVDTLHRAYSKVTEVMHSGRBLGTFRVAFEGQAAYOF 1796

DB 1152 --KILLSODAKTD-----KIVLVN-----THTLKIKNTQFETPAKALTVEV 1195
QY 1797 TDSFVDSGFEDDEGKEYIYEPKLTPLSEISQRLKLYSKPSENNKMIQDSGKNP 1856
DB 1196 NKNKQOL-----QELLNHF-----NAHKAMMDNNHTK 1224
QY 1857 KDLDSKY-AVIOVTV-IPFDEKELOEKTEFER--SHNIRRFEMPEFTQGRQGV 1912
DB 1225 LSPDTKMTAFVSDSLXIPDPFINIKLQDPSIPFVISOGLMKANEMPY-----I 1275
QY 1913 EEOCKRRITLVAIHCPYVYKRI PMVYOHHTDLPLEV-AIDEMSKV-----AELRQIC 1966
DB 1276 QAEITLKLTKLIH-----DINKFCLSAISEFGKEMIPFNVAALRDIE 1318
QY 1967 SSAEDWMTKLOKJGGSVSVQVNAGBLAVARAFLDOTNTRKRPDKKVKLKEVFQFVEA 2026
DB 1319 VOINTKVEIENKVCNETVE-----NTKNIP-KLTKLKQLVPRKV-A 1360
QY 2027 CGQ--ALA VNERLIKEDOLEY---QEEKANRYREMAKELSEI-----MHEQICPLE-- 2072
DB 1361 GGOBOYQVLNKNILSESTSMQOTYERKQKEYFEIYNNVASFKLAFNPQQLQVVERLI 1420
QY 2073 EKTSLVENS 2081
DB 1421 EKFSLPKS 1429

RESULT 3
CENP_HUMAN STANDARD; PRT; 3210 AA.
AC P49544; Q13246; Q13171;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE CENP-F kinetochore protein (Centromere protein F) (Mitosis) (AH antigen).
GN CENPF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=95348175; PubMed=7542657;
RA Liao H., Winkfein R.J., Mack G., Rattner J.B., Yan T.J.;
RT "CENP-F is a protein of the nuclear matrix that assembles onto kinetochores at late G2 and is rapidly degraded after mitosis";
RL J. Cell Biol. 130:507-518(1995).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=95379848; PubMed=7651420;
RA Zhu X., Mancini M.A., Chang K.-H., Liu C.-Y., Chen C.-F., Shan B., Jones D., Yang-Peng T.L., Lee W.-H.;
RT "Characterization of a novel 350-kilodalton nuclear phosphoprotein that is specifically involved in mitotic-phase progression.";
RL Mol. Cell. Biol. 15:5017-5029(1995).
RN (3)
RP SEQUENCE OF 2194-3210 FROM N.A.
RX MEDLINE=9536446; PubMed=7612011;
RA Li Q., Ke Y., Kapp J.A., Fertig N., Medsger T.A. Jr., Joshi H.C.;
RT "A novel cell-cycle-dependent 350-kDa nuclear protein: C-terminal domain sufficient for nuclear localization";
RL Biochem. Biophys. Res. Commun. 212:220-228(1995).
RN (4)
RP CHARACTERIZATION.
RX MEDLINE=95370296; PubMed=7642639;
RA Zhu X., Chang K.-H., He D., Mancini M.A., Brinkley M.R., Lee W.-H.;
RT "The C terminus of mitosis is essential for its nuclear localization, centromere/kinetochore targeting, and dimerization.";
RL J. Biol. Chem. 270:19545-19550(1995).
RN (5)

RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-B reveals interactions with the kinetochore proteins CENP-F and hBUBR1";
RL J. Cell Biol. 143:45-63(1998).
CC -1- FUNCTION: PROBABLY REQUIRED FOR KINETOCHORE FUNCTION. INVOLVED IN CHROMOSOME SEGREGATION DURING MITOSIS. INTERACTS WITH RETINOBLASTOMA PROTEIN (RB), CENP-B AND BUBR1.
CC -1- SUBUNIT: HOMO- OR HETERODIMER.
CC -1- SUBCELLULAR LOCATION: NUCLEAR MATRIX (BUT NOT IN THE NUCLEOLUS), REORGANIZATION TO THE KINETOCHORE/CENTROMERE (CORONAL SURFACE OF THE OUTER PLATE) AND THE SPINDLE DURING MITOSIS.
CC -1- DEVELOPMENTAL STAGE: GRADUALLY ACCUMULATES DURING THE CELL CYCLE.
CC -1- PTM: HYPERPHOSPHORYLATED DURING MITOSIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ch/announce/> or send an email to license@ebi.ebi.ch).
CC -----
DR EMBL; U19769; AAA82889.1; -;
DR EMBL; U30872; AAA82935.1; -;
DR EMBL; U25725; AAA86889.1; -;
DR Genew; HGNC:1857; CENPF.
DR MIM; 600236; -;
DR InterPro; IPR001230; Prenyl_site.
DR KMW Chromosomal protein; Nuclear protein; Centromere; Coiled coil; Mitosis; Phosphorylation; Antigen; Cell cycle; Repeat.
FT DOMAIN 14 197
FT DOMAIN 273 769
FT DOMAIN 823 1328
FT DOMAIN 1642 1746
FT DOMAIN 1862 2987
FT DOMAIN 2207 2568
FT REPEAT 2207 2386
FT REPEAT 2389 2568
FT REPEAT 3015 3032
FT DOMAIN 16 16
FT CONFLICT 250 250
FT CONFLICT 272 272
FT CONFLICT 611 611
FT CONFLICT 1494 1589
FT CONFLICT 1611 1611
FT CONFLICT 1811 1811
FT CONFLICT 2242 2243
FT CONFLICT 2335 2335
FT CONFLICT 2492 2492
FT CONFLICT 2545 2561
SQ SEQUENCE 3210 AA; 367589 MW; 11D83324960E4334 CRC64;
Query Match 1.8%; Score 199; DB 1; Length 3210;
Best local similarity 18.5%; Pred. No. 0.0074;
Matches 424; Conservative 342; Mismatches 828; Indels 700; Gaps 107;
QY 94 CSTVPKAEBAQS-----LFVTECIKTYNSDMLVNYKYEDYSGFERDLPN 140
DB 121 CKSELERQQAQSDVSLNCPNTPKIFTPLPFSQ-----YSGSKYEDLKEKY-----N 172
QY 141 KVKLKDLPVHVYDEVDDDAASISGSKG-----ITKH-----GWL 181
DB 173 K-----EVEERKRLAEAVKALQAKKASQTLPOATMNRRLIARHQASSSVFSWO 220
QY 182 YKGNMNSAIVTMSFKRRFPHLIQLGDSYKFFELKDLQKRPK-----GSIFL 230
DB 221 QEKTPSHSSNSQTPPIRDF-----SASV-----FSGELVTPRRTQIGKRDANSFF 271
QY 221 G-----FLYGVSRNNKVRRAFELKMQDQKSYLLAADSVEVMEWITLTKI--LQLN 282

Db 272 GNSSSPHLLDLKAOQONELRNKINLELR-----LOGHEKEMKQ-----VANKFOELOQ 321
Qy 283 FEAAMOENKNGDSEHDEQSGKLEGGSGLDSTYLPBLAKSR--BAEIKLKS--ESRVKLE 338
Db 322 LEAKKEL-----IEKEV-----LNKCHDELVRTTAQYDQASTYTLAEQTKL 367
Qy 339 YLDPDAQKLDSSA-----BPEVKSFEKFKRILVKNLSPNLQCCVANEBCPTTNYE 394
Db 368 TEDLSQONAESARCSLEOKIKEKEKEFOEL-----SRQ 404
Qy 395 PPRVTLSTFDIKNRKISADFNHDLNHSVRO-----MIATSPALMNGSGPETQSLRGI 450
Db 405 RSFOITDOECIOKAKLTQLOQAKMHNVLQAEIDKLSVXQOLNN-----452
Qy 451 LHEAAMQPKQGFVSTCPHPDIFLARI-EKVLQSGITHC--AEPYKSSSSKVAQY- 507
Db 453 -----LEEFQKUL-----CBAEQAFQASQIKENELRSMEMKENLILKSHSEQKAREVC 503
Qy 508 -----LKNAKQA-----CQRLGQYRMPRMAARLFXDASGNDKNARFSAYRQDSNKLSN 559
Db 504 HLEAELKNIKQCLNQGONFAEMKAKNTSOETLRLQEKIN-----QOENSLTL 553
Qy 560 DMLKLLADPRK-----PERMATPYILGNLDTITINVSPPNYGNSSYIPTKQETCS 614
Db 554 EKULAVADLEKQDCSODLLKKEHNIQLNDKLSKTEKESKALLSLELKKEKEELK 613
Qy 615 KTRPTE-----EVEEFVPCIPKHTOPYTYITNHLVYPR-----YLYKDSQ 655
Db 614 EKEITLFSCKSENEKULIQMESEKENLOSKINLEFLCTQOQKSHENYERVTLEMDIE 673
Qy 656 KSPAKARNIATICEFQDSDEDSQRLKCTYGRPGGFVTRSAFAVLHHQNEFYD--- 712
Db 674 NLSVEIRNLNHNVDLSK-SVEVETQKL-----AYMELOQKAEFQDGH 714
Qy 713 --EIK-----IELPQULHEKHLLLTFFHVSCDNSKSGSTYKRDVVEQVGYMPL- 763
Db 715 QKEIEMCKLCTSQLTQVEDLEHKLQL-----SNEIMDKORCYQODLHAEBSLRDL 767
Qy 764 --KDGWVUTSEOHIPVSAMLPSGYLQYQELGMRHYGPEIKWQDGKPLKISTHLVSTV 821
Db 768 KSKDASLVNEDH-----QSLAFDQOPAMNHSFANILGEQSGMPERSECRLEA--- 818
Qy 822 YQDQDH-----LHNFQYCOQTESGQALGVELVYKLSLHMEGHVMIARPLTLNOL 875
Db 819 ---DQSPKSAIIONVDSLEFLESQKQWNSDLOKQCELVQIKEIEENILMKA--EQM 873
Qy 876 FVLVTRATQBEVAVNTRVLIH--VYAOCHBEGLEBHSYKAYAKAPRYVASEKTYH 933
Db 874 HOSFVAETISQRIKSOEDTSAHONVAE---TJSALENKEKELOLLNDKVEEQAET- 927
Qy 934 BELTKSMTTLKPSADPLTSNKLKYSMFFFDVLKISMAQHLENSKVLLRNQRPASY 993
Db 928 QBLK-----SNHLEDSLKELOLSTLSLEKKEKMSIISLNR-----968
Qy 994 HHALETVMMLPHITQKFRDNPEASKNANSLAVFIKRCFTPMQGFVFKOINNYISCF 1053
Db 969 -----IEELTOENGTLKEINASINQEKMLIQKSESFA-----NYIDER 1007
Qy 1054 AFGDPKTLPEYKFEFLRVVCNHNHYIPLMLPMPFGKRIOR-----YODQLDYSLTD 1106
Db 1008 EKXISISLQYKOEKIL-----LORCEETGNAVEDLSOKKKAQ 1047
Qy 1107 BECFRNLVGLLLEVGTALEFREVRLLAI SYLKUL--IKHSFDRVYASRSHQRIAT 1164
Db 1048 E---KSKKECLINEC-TJSLCENR-----KNLEQLKEAF-----AKEHQEFLTK 1088
Qy 1165 LYLPLFG-----LLIENQRIINVDSPPRVN---AG-----MTVKQESIALPAV 1206
Db 1089 L---AFAEBRNQIMLELETVQOALRSEMTDNQNNKSKSAGGUKOIEIMTLKE-----1138
Qy 1207 NBLVTPQKSGTLDNSLHK--DLGALSGIASPYTT-----SPNINSTRNADSRGS 1255
Db 1139 -----ONKQKQEVNDLLOENQOLMYMKTHQECQNLSEPIPINSVYKERESEBN 1186

Qy 1256 L-----ISTDSGNS-----LPERNEKSNSLDKHQOOSTLGN-S 1288
Db 1187 QCNFKRQMDLBYKEILSDSYNAQOLVQLEAMLRNKKELKLOSEBEKECEQHLEQTRGLE 1246
Qy 1289 VVRCDKLDQSEIKSLMKCPL-----YI-----LKMSDDLFTYMNKASTSELMDFPTI 1337
Db 1247 TSNLOMQSOEISGLKDEIDAEKEYISGPHBELSTQNDNHLQCSLOTWMMKINLEIKI 1306
Qy 1338 SEVCJHQYQWNGKRIARTGMNHR-----LOOLGSLDNLSTFNHSGSDAVL 1387
Db 1307 CEI-----LOAEKELYELNDSRSECTIATRKMAEBVGKLENEKIL---NDSSGL 1356
Qy 1388 HQSLLE-----ANATEVCCTALD-----TJS-----LFTLAF 1415
Db 1357 HGEIYEDLPGEFPGQPHQHPVSLAPLDEBSNYSHELTLSDKEVQMHFAELOEKRLSQS 1416
Qy 1416 KQULIADHGNHMLMKVFDVYLCTLOKQOSETALKNV--TALRS---LYKFPSTYE 1469
Db 1417 EHKILHDO-HCOMSKMBE-----LOTYVDSLAKENLVSTNLNRFQGDVLKEMQGLEE 1470
Qy 1470 GRADWCALCYEILKCSKLSITREASQULLYPLMRNPNFYTGKKSFVRTHLOVITSVS 1529
Db 1471 GLVPSLSSCVB---DSSLSLSDGS-----FYRALRQTDGMSIL-SNLBGAVSAN 1519
Qy 1530 QULIADVIGIGTRFQOGLSIINNCANSDDLKHTSFSSDVLDLTKRIRTVLMATQMKEN 1589
Db 1520 QCSVDEV-----FSSLOTYVDSLAKENLVSTNLNRFQGDVLK-----1558
Qy 1590 ENDPMLVLDQYSLAKSYASTPELRYTLDSMA-----RHHVKNQGLS-----1632
Db 1559 ---EMOGLBEGVLPSLSSCVDPSSLSLSDGSPFYRALBQTDGMSLSSNLGCVSA 1614
Qy 1653 --EAMCVYHNTALVAEYLITREBAVQWMPRLPHSHSACLRRSGRGVFGOGCAFRVITPN 1691
Db 1615 NOCSYDEVFCSLQEBNLTREKTPESAPRAKVGBELSLC-----EVRNQ-----S 1658
Qy 1692 IDEBASMEDEVQMDVHNEEDVLMELLEBOCADGLMKARVEL--IADYKULIIPYERKR 1749
Db 1659 LEKLEEKESQGIKM--NKEI--QELEQ-----LSSRQELDCRKQY---LSEHQ 1704
Qy 1750 DFERLAHL---YDTLHRAVSKYTEVMH-----SGRLLIGTYFRVAFPGQA 1791
Db 1705 WOQKLTSTVLEWESKLALEKKQTEOLSLLEVARLOLOGLDLSRSKSLG-----1753
Qy 1792 AQOFTSETOVBGFPEDED--GKEYIYKEPLTLPSEISORLLKYSDKPGSE---NVK 1846
Db 1754 ---ITDEDAIQGNESCDISKEHTSETTERTPHGHVQ---ICDKAQOQDLNDIE 1803
Qy 1847 MIDSGKYNP-----KDLSKYAYIQVTHVIPFEDEKELOERKTEPERSHNIRFME 1899
Db 1804 KITERGANVLPFGESGEGSDPTNYE-----PRGEDKTOGSSSCEISLSPSGNALVP 1855
Qy 1900 MFTQGTGRQGVQECQKRTITLAIHCFPVYKRIPIWYQHTDNLNIEVALIDMSKV 1959
Db 1856 MDFL-----GNOED-----IH-----NLQRYKETSNEMLRLHVV-IEDRDRKV 1893
Qy 1960 A-----ELRQLCSADEVMIKLOLKQGSVSQVANNAPLAVARAFDINTKRYPNKYKL 2015
Db 1894 ESLNEMKELDSKHLQEOVQMTKIBACTELEKLYVGLKKNSDSE--KLTFSCDHOEL 1952
Qy 2016 LKEVFRQVFAQCALAVN-ERLIKED-----OLEYQEBKANYREMAKELSEIMHQICP 2070
Db 1953 LQRV--ETSQGNLSLEHNAKSRREDIGDVAVANDSKRFLDVENELSRIR-----2004
Qy 2071 LBEKTSVLPNSLHT 2084
Db 2005 -SEKASIEHBALYL 2017

RESULT 4
NM1 YEAST ID NM1 YEAST STANDARD; PRT; 2748 AA.


```

QY 1030 IRRCTFMDRGFVFOIINNYISCFABGDPRTLEFYKFEFLRVVCHNEHYIPLNLPMPFGK 1089
1072 -----LSBSAY-----BELVCKENPDVEFLKEKSAKLGHTVSNSEAYSELE----- 1113
QY 1090 GRQQRQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQDQD 1135
1114 ---KKEOPSLEYLVHAKATNHHLLSDSAYEELVCKENPDVEFLKEKSAKLGHTVSN 1170
QY 1136 -AISVLKN-----LLIKH--SFDDRYASRS--HOARIATLVLPFGLLIENVORINR 1183
1171 EAYSELEKLEOPSLEYLVHAKATNHHLLSDSAYEDLVCKENPDVEFLKEKSAKLGHT 1230
QY 1184 DVSPPFVNAGMTVKDESLLAPVNPVTPQKGS---TLNDSLAKDLGALISGIAEPTYS 1240
1231 VVS-----NEAYSELEKLEOPSLEYLVHAKATNHHLLSDSAYEDLVCK-----KE 1277
QY 1241 TPININVRNADSR--GSLISTDSGNSLPERNS-----EKSNSLDKQOOSTLGNV 1290
1278 NPDMEFLKEKSAKLGHTVSNSEAYSELEKLEOPSLEYLVHAKATNHHLLSDSAYEDLV 1337
QY 1291 RCDKLDQSEIKSLMCFEYLKMSMDALFTYNNKASTSEL---MDFTISEVCLHGFQY 1347
1338 KCKENPDME-----FLKEKSAKLGHTVSNSEAYSELEKLEOPSLEYLVHAKAEOI 1387
QY 1348 MGR-----RYIARTGM--MHARLO-----QSGSLDNLSTFNHSGHSDADVLHOSLEA 1394
1388 OSKISISDPNTLANSPMEDMASKLOKEXOISNDEYIALKNTMEKPEVLLRSKLKGY 1447
QY 1395 NATVECLTALDPLSLFTLAFKQQLADGHNPLMKVNDVYLCEIQKQOSEALK----- 1450
1448 HI-----IDTTY-----NELVSNF--NSEPTLK-----FLBEAKSKGYLLIBP 1484
QY 1451 NWFETALRSLLYKPFSTFYEGRADMACAL--CYELKCKNSKLSIRTEASQLYFLMRNF 1509
1485 NELDL--NRIATTPSK--BEIDNFCQOIGCYAL---DSK-----EYERLAKSL 1526
QY 1510 DYTGKKSFYRTHLQVLIISVQLADVGIIGTRFQOOLSIIINNCAANSDELIKHTSP---- 1565
1527 ENPSRK-----FIEEMALDLVLVDKTEYQ--AMKONASKSKSLISTKVLDPV 1574
QY 1566 -----SSDVKDLTKR-----IRTYLMTAQMCK--EHENDPEN--LVLDQ-----YSLA 1604
1575 TMPAPOLASAKSSLOKRTSLDENELKALGYVALRKENLPMLEKPIVNASKNVDLNIC 1634
QY 1605 KSYASTPELRKTYLDSMARIVHK---NGD-----LSEAMCVNHTALVAEYLTRKEAV 1655
1635 SKFSLVF--LSTEYEDMKRKEHTKIINLGDPSIDFLKECKEYOMLIISKHDEEKQEXAI 1693
QY 1656 QWEPPLLPSSHSACLRSGVAF--ROGCTAFRVITP--NIDEBASME---DVGMQDV 1707
1694 E-----NPGYEFLEKAGALGYELVSEVLEDMKQIMDISPIDVQEKRA 1737
QY 1708 HNEEDVLMELBQCADGLMKAKERY-----ELIADIYKLIIPYIEKRDPERLAH---- 1756
1738 ANREWLLANNEEK--BALQKRIEYPSLTLEIKALAMNMKILVDOIEYDETIRKCNHPTRM 1795
QY 1757 -LYDTLHRAVSKYTEVHSGRRLGTYPFVAFGQAQAOQFTDSETDVSGFEDENDEKEX 1815
1796 ELESCH-----HLNLVLL-----DQNEY 1814
QY 1816 -IYKED--KUTPLPSEISQRLKL-----YSD---KFGSENVKMIQDSGKNPKDLS 1861
1815 STLREPLKORNVEDLINTLSKLNLYAIPTIYQDLGKXENPFYDKXS--LNMKD--- 1869
QY 1862 KXAYIOVTH-----VLPFDEKELQKTEFESNHRIFPMFEMPTQCGKQGGYEEQC 1916
1870 ---YVAISRODYELMVAKYEKQDLVKLSSEKIDHIVPLSRYNLMTVNNYNSPL--SYL 1925
QY 1917 KRTIIP-----TAHGFPPYKRIIPVYOHHTDLNPEV----- 1950
1926 KKAVALNHHILKEDDYKNIILAVSEHPYIYH-----LSEHLLLNKVLVDRNDF 1974

```

```

QY 1951 -----AIDENSKVAAEL-----ROLCSAEVDMIKLOLKQGSVVO 1987
1975 ATMSRIEKPITDFISTKALSWGKILVNESTHKEKNEILLSPDSEFLTMKAQGLIIL- 2033
QY 1988 VNAGPLAARAPLDTNTRKIPDNKVKLKEVFRQPVACQOALAVNERLLKEQOLEQO- 2046
2034 -----SEKEYSE-----LRDQDRNLVLVLEKKAIFISIVEN-IEEQO 2072
QY 2047 -----BEKKNYEMAKELEISEIMHEQICPLEEKTSVLPNSLHIFNAISGPT 2093
2073 LVNTSPCPPIYEDLKYTAQFQMBELCQKPNKLSGAEFRARERIDEGSINTTSS-NSITT 2131
QY 2094 STM 2096
2132 SSM 2134
Db 2132 SSM 2134

RESULT 5
TEGU HSV6U STANDARD; PRT, 2077 AA.
AC P52340;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Large tegument protein.
GN U31 OR HHRE1.
OS Human herpesvirus (type 6 / strain Uganda-1102) (HHV6).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
OX NCBI_TaxId=10370;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9526321; PubMed=7747482;
RA Gompele U.A., Nicholas J., Lawrence G., Jones M., Thomson B.J.,
RA Martin M.E., Efethiou S., Craxton M., Macaulay H.A.;
RT "The DNA sequence of human herpesvirus-6: structure, coding content,
RT and genome evolution."
RL Virology 209:29-51(1995).
CC - FUNCTION: TEGUMENT PROTEIN.
CC - SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 U36,
CC BHV-1 24, EBV BPLF1, HVS-1 64, VZV 22, AND HCMV UL48.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL, X83413; CNA58411.1;
CC DR SEQUENCE 2077 AA, 239946 MW, C1CA4BDC26650511 CRC64;
SQ

Query Match 1.8%; Score 195; DB 1; Length 2077;
Best Local Similarity 17.9%; Pred. No. 0.0064;
Matches 291; Conservative 270; Mismatches 571; Indels 494; Gaps 79;

QY 694 TRSAPAAVLHHQNEDEFDEIKELPTQLHKKH-----HLTLTPFVSCDSSKSTKRD 749
60 TAGTIALMLNNSDSQDY--RLPTBIYKRIHRYGTYGHELSRPFNGTID-----TQKID 111
Db 750 VETQGVSWPLPLKDGKRVTSQHIPIVSAVNPGLGYGDELGMGRHAYEIKWVGSKP 809
112 -----NEVYL-----GLIDFILYKSKN--GPFPAVITIS-V 140
QY 810 LKISTHLSVVTYTDQDQ-----LHNFQYQCKTESGAOL---GNE-----LVKLIK 854
141 LSRALFFLNTLTYLFDHSPTREATAIYQNIIEVVELLTTHGTBEGFYVDASLIFPIE 200
Db 855 -----SLHMEGHVMI-----AFLPTIINLDFVLPRATQEEVAVNVTYII 896
201 TSNLSLSDHDELILKTYKOPDIALDLKFSSTEIHKTKTDIGSQODLVADKTTDLE 260

```

QY 897 HV--VAQCHBEGLESHLSYKYAVAEYVASEYKTHHE--ELTKSMITLIKPSADF-- 950
 DB 261 HAPHRKKNKSHSLLELENDKKKKDSTASYATEVDLIPSEYELSSQFSLPHDKSFPFI 320
 QY 951 LSNKLKSKWFFPDLISMAQHLENSKVLBNQRPPIYHAAVEVVMMLMHPHQ 1010
 DB 321 MKN--FNWITY--LQSPIN--PQOPATPP--LNNRVHLLCOIID 360
 QY 1011 KF--RDNPEASKNANSHLAFVIRKCFPMRGEFVKO--INNYS 1051
 DB 361 VVGVSSTNDSSKQOQOT--IFINYLFPKDPSEVENALAAQENNDLIIFYNNYL- 417
 QY 1052 CFAPDPTLFEYKE--PLRVNC--HENY--IPLNMPFGKRIQRY 1096
 DB 418 C--KTTPTLRLILSKFLAVDNDHKHKEVWVSATTOMFGPKLDIENYL 471
 QY 1097 DLQOLYSLDER--CRN--HPLVGLLREVGTLQFSEVRLAIVLKKLIXSF 1149
 DB 472 KAYVHNPNVKEHFEFLCNKRAEKYKVAVLNKRKEIOAERKNKSFQLSN--F 525
 QY 1150 DRYASRSHQARIATLYPLFGLLIENVQIRINVDVSPPVNAGMTVKDESLALPAVNDL 1209
 DB 526 IDKGE--TPALPIESENVAKHTSDTBEIVRPFIT--ESTELPNIISTL 571
 QY 1210 VTPQKSTLNSLHKLGLAIGIASPYTST--PNINSVRNADSRGSLISTD-- 1260
 DB 572 NNTQOIS--LDKQINELTWTITHTLTKFTKIVODNNNI--AAGFMEVTELCFLAVL 626
 QY 1261 --SGNSLPERNSEKNSLDKHOOSTLNGSVYRCDKLDSEIKSLMCL 1308
 DB 627 VNLVYNIEVLKSHGSLNINTVLLQEVKLYDNTQFLRGSHNNINLSN--FT 677
 QY 1309 YLIKMSDPLATFYNNKASTSELMDFITSEVCLHQFQVVG--KRYIARTGMMHARLO 1364
 DB 678 LSRKMFVU--FNSQKPSD--KASEILAEISLAPSKKATVNIEMIKSQLE 728
 QY 1365 QUGSLDNSIJTFNHS--YGHSDADVLHQSLELANIATEVCLTALDLSLFLAK 1416
 DB 729 ELGKKEISTTENKQFALITKQILGDOELTPYIDFLHLSAVALPNTTVCNHLHLEIQ 788
 QY 1417 NOLLADHGINPLMKKVPDYVLCFLOKHGETALKVFTLRSLIYKFPSTPEGRADMOA 1476
 DB 789 RPDIMATMLDKI--QSLDIYVDMT--NDITYEQTFSTVLVLDLFPNSTERK--T 839
 QY 1477 ALCYELL--KCCNSK--LSISRTESQOLYPLMRNNPDVTKGKSFVRTHL 1522
 DB 840 ALPESVLTIRQAKKCANUKTIDEPDLOFTTMSQONNRQHF--GKK-- 889
 QY 1523 OVIVISQULADVVGIGTRFOOSLSIINNCAUSDRLIHTSFSSDVKDLTKRIRTVL- 1580
 DB 890 --IPLT--MGHIKFLYSOKI--TTEBKWMIQRAKTVVIT 923
 QY 1581 --MATP--OMKE--HENDPEMLVDLOYSLAK-- 1605
 DB 924 SPEELTAFATAPTKALQOTCKPDLDKALQRMHEBOMKQJANDKXKILITISTLEKRLT 983
 QY 1606 --SYASTP--ELKRTWL--DSMARITHVKNGLSEAAVCYVATVLMVAY 1648
 DB 984 DILLILKQGFSSLETMLNLLETFLKQLODNNVNIHFTALLPVLKDIETTISKISIV 1043
 QY 1649 LTRKAVOWEPPLLPKSHS--ACLRSRSGV--ROGCTAFR 1686
 DB 1044 I--EKILIKTLPNPOKSKSEOKYTPILSPISKFKTTCFEDVTEIEQYOKSITPLX 1100
 QY 1687 VITPNIDEASMEVDGMDVAFNEVDVLMEL--LEQCADGIMKAKERYELIADIYK 1739
 DB 1101 KATSTNKHTRLSHSYGQELMLYERITELKKTNNKKEQJSEKVALAEK--K 1152
 QY 1740 LIIPIYKRDPERLANHLVDTLHRAVSKYTEVWHSGRLLIGTYFRVAFQAOYQPTDS 1799
 DB 1153 ILLSQDADKT--KIYLVAN--THLIKAKTQFKETAPAKALTYEVANNK 1198
 QY 1800 ETDVEGFEDDEQKEYIYKPLTPLSEISORLLKLYSDKFSSENVKMIQDSGKVNPKDL 1859

DB 1199 ENQJ-----OELNHF-----NALLKAKMPOHNTKLSF 1227
 QY 1860 DSKY-AVIQWTV-IPFDEKELQERKTEFER--SHNIRRFMEFPQTGKQGVBEQ 1915
 DB 1228 DTKMTAFVSDSLYIPDITNKLQDPISDPKVIQSOLMKNKTNEMFY-----IQNE 1278
 QY 1916 CKRRTILTAIHCFPVYKRIIPVMTQHTDNLPIEV-AIDENSKV-----AELROLCSGA 1969
 DB 1279 ITLKMLTLLIH-----DINKFCLSAISEGKEAIPNNYALRLBEYQI 1321
 QY 1970 EVDMIKLOLKQGSVVOVNAGPLAVARAFLDNTKAYPNNKYLKVEPRQVYACQ 2029
 DB 1322 NTKVYEIENKVIQNEVE-----NTKNIP-KLTKLKLQDLPKRY-ACGQ 1363
 QY 2030 --ALAVNERLIKEDQLY--QEBKANYREMAKELSEI-----MHSQICPLE--EKT 2075
 DB 1364 EOYOTLMKRLITSETSMOQTEKEQLKREYFEIVNVASFPLAFNPPOOLQONVERLIEKP 1423
 QY 2076 SVLPNS 2081
 DB 1424 KSLPKS 1429
 RESULT 6
 G160 MOUSE STANDARD; PRT; 1325 AA.
 ID G160 MOUSE
 AC P55937;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Golgin-160 (Male-enhanced antigen-2) (MBA-2).
 GN GOLG3 OR MEA2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CD-1; TISSUE=Testis;
 RX MEDLINE=97217683; PubMed=9063644;
 RA Kondo M., Suto S.,
 RT "Cloning and molecular characterization of cDNA encoding a mouse
 RT male-enhanced antigen-2 (Mea-2): a putative family of the Golgi
 RT autocaligin".
 RL DNA Seq. 7:71-82(1997).
 CC -1- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN SPERMATOGENESIS AND/OR
 CC TESTIS DEVELOPMENT. PROBABLY IDENTICAL WITH THE SEROLOGICALLY
 CC DETECTABLE MALE ANTIGEN (SDM).
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN TESTIS. TRANSCRIPTS CAN BE
 CC FOUND IN SPERMATIDS DURING SPERMATOGENESIS. NO EXPRESSION IN
 CC LEYDIG CELLS, SPERMATOCONIA, OR SPERMATOCYTES.
 CC -1- SIMILARITY: HIGH, TO HUMAN GOLGIN-160.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MEA-1, MEA-19 OR MEA-30 IS THE
 CC INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 CC DR EMBL, D78270; BAA19612.1; -.
 CC DR MGI, D78270; BAA19612.1; -.
 CC DR MGI, MGI:96958; Golga3.
 CC KW Spermatogenesis; Developmental protein.
 CC FT DOMAIN 201 204 POLY-ALA.
 CC SEQUENCE 1325 AA; 149880 MW; 3230636962C687B0 CRC64;
 Query Match 1.8%; Score 193; DB 1; Length 1325;
 Best Local Similarity 19.2%; Pred. No. 0.0042;
 Matches 191; Conservative 172; Mismatches 389; Indels 242; Gaps 46;

```

1215 GSTLD---NSIMKDLGAIISGAPYTSTPNINVRNADSRGSLISTDSGNSLPERNGE 1271
99 GSTADAVGSLKKS---RSSTSVSEVSPSESTDNRESASMTGDSSEADGN---ESDS 153
1272 KSNLSLDKHOOSSTLGN-----SVRCDKLDQ--SEIKSLIMCELY----- 1309
154 SHSLSARGACVGLVGNVGPSTAYMVWGOEISAEALGPPSIDVLAQAAAHODONQBA 213
1310 ---ILKMSPDALFTYNNKASTSELMPF-----TISEVCLHQFOVWKGKVIAR 1355
214 NGEVRNRDRSICSSVSMESSELPEDELIQILKORRLEGVAFSLASQALQEK----- 269
1356 TGMWHAARLO-QLGSLDNLSTFNHSGH-----SPADVLHQSILBANIATEVCLTA 1404
270 ---AELQAQLAALSTRLOAQVHEHSHSQODSSSEVDLTQKSCWDLGAMTDLQGM 324
1405 LDTLSLFTLAFKNOI--LADHGNPLMKVFDVYLCLQKHQSESTALKVNFALRSILYF- 1462
325 LEKNAASLASNNDLQVABEQYRLMAKVEDMORNLTKONTVHDLRQOMTALQSOLOQV 384
1463 ---FPGTYGGRADMCAALCYELKCNKSLSTRBASQLYFLRRNPFDTYGRKS 1516
385 QLERTTLTSKLQASQAEI-----TSLQHARQWYQOQLTLAQBARYRLQGENA 431
1517 FVRT-----HIQV-IISVQLIADVIGIGTRFOQSLSIINNANSDEL-IKHIS 1564
432 HIQVGMTOAGLEHLKLENVSLSHQTLER-----QHRI-----KEKERIANQLOS 478
1565 PSSDVQDLTKRIRTVLMAITQKMEHNDPMLVDLYSLAKSYASTPELRKTWLDSMA- 1622
479 IEADMDOEAFAVQIREAKTWEE-----DLORLREEFEGEREQLOKV-ADAAAL 528
1623 ---RIHVKNKGLSEAMCVVHTALVAEYLTRKAVQVMPRLPHSHSCLARRSG 1675
529 EQGLEOVKTLFRDQDLALQOEHD-VIKULTSTQELQAQGLDLDH----- 578
1676 GVFRQCTAFRTVPINDE---EASNMEDVGMQDVHF--NEDVIMLEBQCADGLWKAER 1730
579 ---TRYDELQARLELQREADSRDA---HFLQNEKIVLEVALQSA---KSNK 623
1731 YELIADYKLIPIYKRRDPFERLAHLYTLHRAYS--KTEVMSGRRLIGTYFRVAFRG 1789
624 BEIDRGARRL-----BEDTETSGLLEQRLDRLAVKSNQVHLCQETATLRKQM- 672
1790 QAQVQFTSETDVEGFEDDEGEKIYKEPKLT-----PLSEISORILKLYSDK--RG 1841
673 QKXKEQFVQOKWVEAYRRDAYSKDQILNELKATKRRLDSEMEKELQOBILKQGEKTYE 732
1842 SENVKMIODSGKNPK--DLDSKYAIVQVTHVPFPEDEKLEORKTEF-----ERS 1890
733 VHSRRLQKMSLVHQMALEHGLSQVQKER-----DEMEIHQSLKPFKEQMIALTEAN 787
1891 HNIRPFEMPTQTSKRGQVGEQKRRTI--LTAHCFPVYKRIPIVMTQHTDLP 1947
788 ETLKXQIEEL---QOEAKKAITQOKMKRLGSDLTS-----AQEMTKHKAY----- 833
1946 IEVAIDEMKVAELQOLSSAEVMDIKYOLKQSGSVQVNAAPLAPARPLDNTNRK 2007
834 -ENAVASILSRLOE--ALASKKATDELNLQRPQ---STGSSDPVLAKIRALELEON 887
2008 YPDNKKLLEKVEVQFVEACGALANVERLIKEDOLEYOEMKAN--YREMAKELS----- 2061
888 VQGSKILREKL-QEYITMTSOLERS-----REKYLEDELQESGFRKRIKRLAESNK 942
2062 ---SIMEFO-----ICPLEKTSVLPNSL 2082
943 KLALEHERGKLTGLGOSNAALREHNSILETLAL 976

```

RESULT 7
DHC_CABEL
ID DYC_CABEL STANDARD; PRT; 4568 AA.

```

AC Q19020;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dynein heavy chain, cytosolic (DYHC).
GN DHC-1 OR T21E12.4.
OC Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=96114101; PubMed=8674131;
RA Lye R.J., Wilson R.K., Waterston R.H.;
RT "Genomic structure of a cytoplasmic dynein heavy chain gene from the
RT nematode Caenorhabditis elegans."
RL Cell Motil. Cytoskeleton 32:26-36(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Du Z., Maggi L.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: DYNEIN HAS ATPASE ACTIVITY. CYTOPLASMIC DYNEIN ACTS AS A
CC MOTOR FOR THE INTRACELLULAR RETROGRADE MOTILITY OF VESICLES AND
CC ORGANELLES ALONG MICROTUBULES.
CC - SUBUNIT: CONSISTS OF AT LEAST TWO HEAVY CHAINS AND A NUMBER OF
CC INTERMEDIATE AND LIGHT CHAINS.
CC - SUBCELLULAR LOCATION: Cytoplasmic.
CC - SIMILARITY: BELONGS TO THE DYNEIN HEAVY CHAIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL, L33260; AAC37251.1; -.
CC DR EMBL, U80440; AAK21472.1; -.
CC DR WormPeP, T21E12.4; CE23997.
CC DR InterPro, IPR003593; AAA_ATPase.
CC DR InterPro, IPR004273; Dynein_heavy.
CC DR Pfam, PF03028; Dynein_heavy; 1.
CC DR SMART, SM00382; AAA; 1.
CC DR Motor protein; Microtubules; Dynein; ATP-binding; Coiled coil.
CC
CC FT DOMAIN 587 652 COILED COIL (POTENTIAL).
CC FT DOMAIN 814 844 COILED COIL (POTENTIAL).
CC FT DOMAIN 1241 1274 COILED COIL (POTENTIAL).
CC FT DOMAIN 1324 1340 COILED COIL (POTENTIAL).
CC FT DOMAIN 1559 1591 COILED COIL (POTENTIAL).
CC FT DOMAIN 1964 1992 MICROTUBULE-BINDING (POTENTIAL).
CC FT DOMAIN 3132 3229 COILED COIL (POTENTIAL).
CC FT DOMAIN 3339 3432 COILED COIL (POTENTIAL).
CC FT DOMAIN 3707 3739 COILED COIL (POTENTIAL).
CC FT DOMAIN 4359 4386 COILED COIL (POTENTIAL).
CC FT NP_BIND 1865 1872 ATP (POTENTIAL).
CC FT NP_BIND 2163 2170 ATP (POTENTIAL).
CC FT NP_BIND 2537 2544 ATP (POTENTIAL).
CC FT NP_BIND 2860 2867 ATP (POTENTIAL).
CC SQ SEQUENCE 4568 AA; 521568 MW; 028B52684F381676 CRC64;

```

Query Match 1.7%; Score 190.5; DB 1; Length 4568;
Best Local Similarity 18.0%; Pred. No. 0.039;
Matches 409; Conservative 319; Mismatches 798; Indels 741; Gaps 114;

Db 66 NVIVDVRVAREQGOOPAGESGSEESADATFOVHDLFTMTDRCQAMFVKOS-----N 119
Qy 204 LIQ-----LGDGS--YKPEFLKDLQKEKSGIPLGFLYGVSPNNKYRFAF 248
Db 120 VIEBAEKIATQVSAPFLNGSAMQOHLFMSRLNLNYCKS-ITG-QSGREBERODKLAFLV 178
Qy 249 ELKMODKSYLLAADSEVEMEWITLILKILQINFAMQEKRN-----GSHDEDEGS 302
Db 179 QCFTEAEALHLQONIDIPEINILVNOHIIIDALEQAGKENBRAKIEDGLDVED--AN 236
Qy 303 KLEGGSSGIDSYLPELAKAR--EAEIKLSSESIVKFLYDPRDQKLDPSASBEVKSFE 361
Db 237 FNLALQSGCNRWVKEIRKVTQLERDPSGSGISLOEMTFWNLBEALLKIS-----OK 287
Qy 362 KFKGRI-----LVKCNDSLFPNLQCCVABNEBEGPTTVEPEFVTLSPD-----IKYRKIS 412
Db 288 RGESEVTLJLEALKG-----KRFHATVG-FBDDNGLKQCLAVY 325
Qy 413 ADFHVDLNFVSROMI-ATTSPALMGSGPEQTOSALRGI-LHEAM--QYRKQIGSVT 467
Db 326 QVNTLMKIFPLSELVASATDVPKLMH-----AVVGIFLHRKLRSTKFLORA---- 373
Qy 468 CPHPOLVARIKLEKLOGSITHCABEPMYSSSDSKVAQVLRKAKACQRLQGYRPMFAM 527
Db 374 ----LRLVEAISRDLSQLKLVLSY--NLKRTPIAE--FNEIMSQC----- 413
Qy 528 AARTLFPKASGMLDKKARFSAIYRQDSNKLSDNDMLK-----LADPR 570
Db 414 ----ALFSKDDDEYDK--FIALLR-DINKKGRDDPSKLSIKYTAHVKLETRLMQLOLR 466
Qy 571 KEKRAKT-----PVILGN-----LDI-----TIDNVSSDP 597
Db 467 KQHEGRTYIERKRVGNGSREBEQMLIDSSEKSPDEQVDIAVEFLKNDVFLDVPD 526
Qy 598 NVVNSYITKQKET--CSKTPITEVEBEFVPCIKHQPTIYI--NHLVYPRY--- 649
Db 527 AEMNA--FKRYEQOIGVETAITTRIKLSQLESSNSNEMFSIFERYALFPRIRIGA 582
Qy 650 ----LKVD--SOKSFAKARNIAICIEFKDDEBDSQPKCIYGRPGGVFPR 695
Db 583 IYEQTRLINRYKEDINELQARFTKRG-----EQYK-IMQVYGLRPPSA 627
Qy 696 SAPPAVLHHQNPEDYDEIKIELPQOLMEKHLLLPFHVSCDSSKSGSTKRDVETQV 755
Db 628 KIMJIRNRYEROLORYKRVEDVLGQ-----WENHVDGQLADGDNFV 672
Qy 756 GYSWRLPKDGRVYVTEQHIPIVSAN--LPSGILGYELGKGRHYGPEIKVNDGKFL-IX 812
Db 673 KLNTQPMFDEM--VESVQ--SQMWTLPNKILITVDV-----QVDRMQLQK 715
Qy 813 ISTHLVSTV-YTQDOHLH-----NPFQYCOQTESGAQALGENELVKYLSLHA-- 858
Db 716 IYHSSSEVLKYKEVSHLKSMMGRVPLKYIVMAHQANQMPSATSL-IEAKRFPASVNAAL 774
Qy 859 --MEG--HVMIAFLPTILNQLR-----VLTRAQEBVAVNVR 893
Db 775 ASVQGVDSLASYKQDIONQLEGATLGWDSYKVDQYQKLAETVTVYGERBELLNVVR 834
Qy 894 VI--IHVAQCH-----BEGLESHRASY 915
Db 835 IYNADINLTKSGRYDKETIENLITSIQGVDSQSLGNYNLQMWNTLPROJETTILARV 894
Qy 916 KYAYKABPVASBYKYVTHEELTK-----SWTTLKPSADFLTNKIL 957
Db 895 EDHAKRWTLVFSQSEEV-BELREQOVVLPVKNVVDLCKMTAOTLYISSTR-ETREKIL 952
Qy 958 K--YSWFFPDVLIKSNAQHLIENSQVLLRNQRF-PASYHHAVETVNNMLMHIQOFD 1014
Db 953 EOLYEMHSV-----CTAOWRISGKRFQVMNBEIEBETHN-----ILN-VMPEGOACLEK 1002
Qy 1015 NPEASGNANHSIAVFIKRCFTF-----MREGFVKQKINNYISCFAPAGDKITLPEYFEBL 1069
Db 1003 AYDCVNGIMSIDLEBYLSEWLSYOSLWVLAQOELFEMLGISLSKM-----KITMEIRKG-- 1056

Qy 1070 RVVCHNEHYIPLNLPMPGKRIQRYODLQDYSLTDEFCRNHPLV--GLLREVTALQ 1127
Db 1057 RLVEPTQOTRKVIFPVSVBYGKAQ--QKILFRY--DYMHKEMLVKSGAV--VGDEMQ 1107
Qy 1128 EPREVRLAISVLKXLLIKHSPD-----DRYASRSHQ 1159
Db 1108 KF-----FNSVSKRWVLEQTSVDSGSDTIGLISFVOSLKKQTKSGDAVDLYRSSQRL 1163
Qy 1160 ARIATLVIPLRGLLENNOR-----INRVDV-----PPPVNMGVTKOESLLAPVN 1207
Db 1164 LMQRYQFPQOMVSENVGEGMSAFETILSLDASIQOMMNLQTKPAQOEBLVEKRTVE 1223
Qy 1208 PLVTPQKSTLDSNLH-KDLGALSIGIASPYTSTPININSVANA-----DSRGSLLISTDG 1262
Db 1224 TLTENMKKRPVGAQRPOEALNVITAFPAKUNKLTHEENKORAKAVALDLSASAPBEG 1283
Qy 1263 NSLPERNSEKS-----NSLDHQOOSTLGSVVRCDKLDOS-----EIKSL 1303
Db 1284 DKLTATEELAAKQVWKALQPVVYTGIDAEKKTLS--VQPRKIRBSLDELAMQLOK 1340
Qy 1304 -LMCFYLILKMSDALFTYNNKASTSELMDEFTISEVCLQFOYMGKRYIARTGMHAR 1362
Db 1341 PVKCRTYASYEHVKOMLHYGKMMNLVABLSEALKENHMO--MMKEMRYNNLSDLT 1397
Qy 1363 LOQLGSLNLSLTFNHSYGHSDADVLHQ-----LLEANATEVCTALD-----TL 1408
Db 1398 LGQVW-----DADILRHETIKILLVVAQGEALAEFLREMRVQWNTVEV 1442
Qy 1409 SLFTLAPXQUL-----LADGHNPLMKRVFVYLLCFLQKQSETALKNVETAL-- 1456
Db 1443 ELVYNQNTRLIKGMDLFPNKLEHQNSLSAMKISPPYKQFEESQSDERKLNKINAFD 1502
Qy 1457 ----RSLIYFPPTFEGRADMCALCYELIKCNSKLSIRFESQLLYFLMRNPD 1510
Db 1503 WVIDQRRWVYL--EGLSSGAEISTLPPF--SSRPATITDVTALM-- 1546
Qy 1511 YTGKSPRTHLQVIVSVOLADVYIGGT--RQOQSLIINN- 1553
Db 1547 --KR-----VAASPRILDVVMQGAORLLERLADMLAKIQXALGEYBERBSF 1593
Qy 1554 ----ANSRDLIKHTSFSSDVYDLTKIRITVLMATAQKHEHDEML-----VDL 1599
Db 1594 PRFVFGDEDLLEIGNSKIDITRIQKHLKMPAGITADINEEDRSITAFHSREGEKYL 1653
Qy 1600 QYSLAKSASTPELR-KTWLDSMARIHVKNQDLSLAACVYHVTALVAYELTRKAYQME 1658
Db 1654 ----VKIVSTDVAINDLQAL--BAEMKHTLARQLAASLTHFSKQNIQTMTTDDYVEM- 1706
Qy 1659 PELLPHSHSACLRBSRGVFRQGTAFRVTIPNI--DE-EASNMEDVGMODVHFNEDVL 1714
Db 1707 ----LDKRPAYI-----TLTAEIMCMEKETTILADGKAENV--EQAV 1744
Qy 1715 MELLEQCADGLWKAERYELIADIKYLLIPIYKRRDEFLRLAHLVDTLHRAVSKTEVMH- 1773
Db 1745 VKTIELLADSVLKEBP-----PI--RRKQMEAL-----ITELVAK 1777
Qy 1774 --SGRRLIGYFRVA-FRG--QAAQOYQFTDSETV-----EGFPEDEDSGYLYK 1818
Db 1778 RDTCKLVSMMKIRANDFGMLQCMFYPDPKQVDPVRCVVVMANSQFFY--GEYIGI 1834
Qy 1819 EPKL--TPLSB-----ISQRL-LKLYSDKFG-----SENVYMIQDSKVPNKDLSKYA 1864
Db 1835 QERLVRTPLDRCYLTMTQALHSRIGSPFPGAGTKTSYKAL----- 1878
Qy 1865 YIQVTHVLPFDEKELQERKTEPERSHNIIRPFM-----EMPTQYKSGQGVBEQCR 1918
Db 1879 ----GQOLGRFVLVFNCDFTFDFQAMGRILVGL--CO- 1909
Qy 1919 RTILTAHCFPVYKRIIVMVOHTDLNPIR-----VAIDENSKYVALEROLCSSAEVDMI 1974
Db 1910 --VGAMGCF-----DEFNRLBERMLSAVSQOIQTIQAVRAAGDMSVDLV 1952

OY 1975 KLOJLQGVSVQVGNAGPLAVARAFLDITTKRYPPNKKVLLKEVER 2021
 Db 1953 GKRLVNSNIGIFITMNPVSGRSNL-----PDN-----LKQLPR 1987
 RESULT 8
 ID RBP1_PLAVB STANDARD; PRT; 2869 AA.
 AC 000758;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Reticulocyte binding protein 1 precursor.
 GN RBP1.
 OS Plasmodium vivax (strain Belen).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92315338; PubMed=1617731;
 RA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;
 RT "A reticulocyte-binding protein complex of Plasmodium vivax
 RT merozoites";
 RL Cell 69:1213-1226(1992).
 CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
 CC -1- SUBUNIT: HOMODIMER (POTENTIAL).
 CC -1- SUBCELLULAR LOCATION: Membrane-bound.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M88097; AAA29743.1; -
 KW Malaria; Receptor; Signal; Transmembrane.
 FT SIGNAL 1 17 POTENTIAL.
 FT CHAIN 18 2869 RETICULOCYTE BINDING PROTEIN 1.
 FT DOMAIN 18 2807 EXTRACELLULAR.
 FT TRANSMEM 2808 2826 POTENTIAL.
 FT DOMAIN 2827 2869 CYTOSOLASMIC.
 FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
 SO SEQUENCE 2869 AA; 330213 MW; 89DBE4422058BCFF CRC64;
 Query Match 1.7%; Score 187.5; DB 1; Length 2869;
 Best Local Similarity 17.1%; Pred. No. 0.028;
 Matches 433; Conservative 408; Mismatches 885; Indels 809; Gaps 118;
 OY 34 SEVVRGSVLLAKKLEPLDYENVIVOKTQIANDC--LREMLFFYDDPQRIARQGR 91
 Db 480 SKVKSNEELS--TIIDLGKSAVLAQESTPQDEBCNKIKTEAEKVDADAEDICEKNEDI 537
 OY 92 YICSTYPAKAE-----EAGSLFVTECIKTY-----NSDWHLVNYK-----YE 129
 Db 538 YV--EIPSEDEITIDKINDLODLI--DQMEKYDVIANNSEFISRIKRIYENLKETEYE 593
 OY 130 DVSGEFRQLPNKVVKLDKLPVHYVEVDEVDKEDASLSGOKGKITKGWLYKGNMNS- 188
 Db 594 TELNDIGKLENDTSKVNFLMQIRIKINTEKTIKIDESLQIVTEK-----FYKEILDSEK 644
 OY 189 -AIVTMRSGFKRFFPHLIQLGDSYKFEFLKDLQKPKSIFLGLVLYGVSPNNKVRRA 247
 Db 645 EKYTELKIEFKSVTEINRLQDG---ESAPDLHEQIKEL----- 682
 OY 248 FELKMDKSSYL-----LAADSEVEMEWITTLNKLQNLFE--AMQEKRNDSH----- 296
 Db 683 --DKMAKQHYHLELLSLKSKSVYFTENNELNLTASYDNMBGFSKKKEADADIDIALYN 740
 OY 297 -----EDDEQSKLEGGSGGL-----DSYLPRLAKSAR--AEIKLKESRVLKFLY 339

Db 741 SVYREDINALIBEVEKFTJENKSTLEMLKDEMEBEKLODAKETAKLNFVSDDKLTDVY 800
 OY 340 LDPDAQ-----KLDSSABPEVYKSEBEFKRIYLKCDL--SFNIQ--CCVAEN 385
 Db 801 TKMSAEVNTNAGIKKEIAQKQFENVHKMKKEKSDAFSTKEALQNSMOQYQDEGAIEGH 860
 OY 386 EGGPTTNVPEFVTLSTL--FDIKYNRKISADPFVDLNHFSVROMIATTSPALNAGSGPETQ 444
 Db 861 KQNBSEKEEYFKMSEVEEDLSRETEBOEYTKHNNSSRRK--GEISAEITN----- 911
 OY 445 SALRGILHEAANQYKQGI---PSVTCBPHPIDFLVANI--EKVLOGSITHCAPY----- 494
 Db 912 --REVINKIESQLNVYGVIEKYFSLIGQNEVSTAKAKETIVSDSLRDKIDQYETEYFK 969
 OY 495 -----MSSDSSKYAQAQVLYAK-----QACQRLGQYRMP- 524
 Db 970 EKTSAVENTVSTIGLSKIDSLKRLNGSINNCKRYNTDILRSKITLREEVGKEMPK 1029
 OY 525 -----FAMAAATLFPDASGNLDKNARFSAIYRQDSNKLNDMLKTLADFR----- 570
 Db 1030 RQDKCGENTTALLKSL--RDKMGKINEKLNDRGLMSLDTK--EDLKFYSSESKITHL 1085
 OY 571 -KPEKMAKLPV-----ILGNLDITIDNVSSDFPN-----YVNS--SYLEPT 608
 Db 1086 SKDQKGPQDPLNRIDEMEDIKRDVDLNVVQVISENKVTLFKNNSVTYIEMHSHINTV 1145
 OY 609 QPE--TCSKTPITFEVE----- 1145
 Db 1146 AHGITSNKEILKSVEVEDKLNLEQNEVDYKVPKNENKQLEAIRGSMKKEVINHG 1205
 OY 633 TQPYTYTNHLYVVPKYK-----YDSQSFAPKARNAIACIEFK 671
 Db 1206 VSEMTQDSTANTLKSNAKKNENHDLBELNKTQGMQDIYKLLKIBELKEGVNLEK 1265
 OY 672 DSDEE---DSQPLKCIYG-----RPGGVFTRSAP----- 698
 Db 1266 DANERANKKPEPEPHNIIGHVLERITVEKDKAGKVEEMNSLKTIEKILQETSDSQNE 1325
 OY 699 ---AAVLHHQNPPEYDEI--KTELPMQLEKHHLLLTFFHV-----SCDNS 740
 Db 1326 LVTTITTHLENAKYEDVIRKNEEDSITQLEKAKSLTLEBMKLVQOVNNNLSAIOG 1385
 OY 741 SKGSTKK---RDVVEOVGYSMPLK-----DGRVVTSE--QHIPVSA 779
 Db 1386 NAGISEKLEKGVLELLISTNYSISILEYVKNKSSSEVRFQNLNGERTKAGEEKNSA 1445
 OY 780 NLPSEY-----LGVOELMGGRHYGPEIKWVDGK--PLIKISTHLVSTVYTQDQHLH 829
 Db 1446 RLAEAEKLEKQIVKLDVSDI-----DDKYKIEGIRKELIKMK-----ESALTFWESSE 1495
 OY 830 NFOYCKK-----TESGAQLAGNELVYKLYSLHMEGHV 863
 Db 1496 KFKQKSSHMEKNAEKKKIEYLKKNNGDGKANITDSQMEBEVGNVSAEHNHFTVEQV 1555
 OY 864 MIAFLPTILNQLFRLTRATQOEVAVNTR-----VIHVAQCHEEGLESILRSYV 915
 Db 1556 -----DKTAFCEISIVAYTTQNDLNFENSLMEKVEYKCKCEKNDKAE----- 1596
 OY 916 KYAYAAEPY-----VASEKYVHEELTKSMTTI--LKPSAD 949
 Db 1597 KYSATLKPYDGRKIRAVSENERKISLEKAKAVEKESQSLNDVSTKSLQIDNCRQOLD 1656
 OY 950 FLTSN--KLKYSWFFPVULIKSMAQH-----IENS--YKYLNRKRPASVYHNAVET 999
 Db 1657 SVLSNIGVQVAAQVFDSPADSKMSKSVLPISLGAELKSLDYKAKA-----SEKNELET 1711
 OY 1000 VNNMLM-----PHITQKFPDNPDEASGNANSHLAV 1028
 Db 1712 VQNEKSRINVEGSLTIDDKITDLENDLLKMKQYEGGLQIKENADKK--SNFEL-- 1768
 OY 1029 FIKRCFTMDRGFVYKQIANNYISCFAPGDPKT--LFEYKFEFLRVVCHENY--IPLN--LP 1084

```

Db 1769 -----VSGEINALL-----DPSIFIKLKKEYDMTGDKYGVKXNEIH 1809
Qy 1085 MPFGK--GRIOYQDQLDYSLTDEFCRNHFLVGLLREVWTAJOF--REVRLI----- 1135
Db 1810 GEFYTSYNLIETHLSATDYSTFEKAQS-----LRRLAEKEEHLRRREEAIFLPLN 1862
Qy 1136 -----AISVLKNLKHSPDDRYSRSHQARIATLYLPLFGLLENNORINVRDVSPPP 1189
Db 1863 DIKVEYSLTKLKEKMEKVSAYEGMKRDHTS-----VQLVQDMKTI----- 1904
Qy 1190 VNAGMTVQESLALPVPNPLVPQKSTIDNLSLHND-----LGAIGISPYTTSTP 1242
Db 1905 VELKTLNDISCCSVLNNVSVIVK--KVESHADVRDANSWESMTLNLFLSDBA 1962
Qy 1243 NINSVR--NADSRGSLSTD-----SGNSLPERNSKSN--LDHQOSSTJGNGV 1289
Db 1963 KISSGMEFPAEMK-SNFKTDLLEIRISVINSNELKKEQDSNDVIOKERESEQLAK-- 2019
Qy 1290 VRCDKLQGEIKSLMCPYLILKMSMDALFTYWNKASTGEELMDFPTISEVCLHOPQYNG 1349
Db 2020 -----DATDIYVNIK-----LKNFNEKLEBAKNK-----BEVSEKVRBAL 2056
Qy 1350 KVIYATATG-----MHARL---QULGSIDNSLTFPHSGHDADVLHOSLEANATEV 1400
Db 2057 KRLSOVEGRCHRENFRLLDTELENKQVVTIYRDKISERESGLQEMENEMTYSN- 2115
Qy 1401 CTALDPLSLFTLAFKNOCLADHGNPLMKKVDVYLCLFQKHQSESTALKNV---FTALR 1457
Db 2116 SITOLEGIYVSAGESKEDI-----EKL-----ERSHEEMENISEKSTID 2155
Qy 1458 SLIYKPSSTFYE-----GRADMCALCYEILKCNSSKLSIRTE--ASQLLYELMBN 1507
Db 2156 SKYIEMNSTIDELYKLGKNCQAMWISLISYANMKTSKLIMINKENKTEKVDVIKON 2215
Qy 1508 NRDYTG---KSGFVATHQVIVSQQL--ADVAGIGGIRRO-QSLSIINNC----- 1553
Db 2216 SSSDTGVEYETLKGFGYSGKL-TRESSASEIYQMDLYSVNPAKHEKESLNAIRIDKELYL 2274
Qy 1554 -ANSD-----RLKHTSPSSDVQULTKRIRFVLMATAQOMKEHNDP--EMLVDLQY 1601
Db 2275 HONSDSIYEGVQNMALYDKLNEKREMDLYRISETKLYKOMESHSTDFVEKPEMLK 2334
Qy 1602 SLAKSYASTPEBLRKTWLDMSMARLHVXNGDLSBAACUYVATLVAEYLTRKEAVOEPPL 1661
Db 2335 GNNETN-----NKSLEKEKKLKSVND-----HMHSSEAMT--KNGLKATPSS 2376
Qy 1662 LSHSHSACLRBSRGVFRGGCTAFRYITPNIDBSASMEDVGQDVHNEVDLMELLBOC 1721
Db 2377 VONINN-----IYGVIEAEVKTLEHID-RDYGDMVOIYEEHKKO- 2414
Qy 1722 ADGLMKAIHYELIADLYKLI--IPYERKRDPERLAHLYDTLHRA--YSKVTEVMHSG 1775
Db 2415 -----FSLIDBTNALMDIDIEFKKENNNLEAVNETITHRVNDYIEKTLNKLVOA 2465
Qy 1776 RRLGLTYFVAFFGQAQYQFTDSESTDVEGFEDDEGKEYIY-----XE 1819
Db 2466 K-----TEBOILENIKQNDMLNIFLKVYSIIEFENYKKE 2505
Qy 1820 PKLTPLSISORLLKYSDKFGSENNKMIQDSGVKNPKULDSKYAVIQTTHVIPFDEKE 1879
Db 2506 SILNDIYE-QERLLKI-----GEHL-----DEIKRNVITLSEYEDQEMEM--SGN 2550
Qy 1880 LQERKTEPERSHNIRFEMFPQTQGXKQGVGEBQCKRITLTAHCPRYKCR--IP 1936
Db 2551 LBEKSKSMNYTSI--YELERANEENRRAKQIKD--DTLINSV-LEAALQKRDMDA 2604
Qy 1937 VNYQHTDNLPIEVALIDEMSKVAELRQCSSAEVMDIXLOLQCSVSQVNAAGPLAYA 1996
Db 2605 IFSQMSADNPNEYKSAE--KWNENANEIIROLEVGLRIGQVQDSSEI----- 2652
Qy 1997 RAFPDDTNKRYRPNKVVKLKEVFRQVBAAGALAVNERLLKEDOLEYQOEMKANYRBM 2056
Db 2653 ---LSEWNSKGAIEK-----EKTAALATISNNRE-----EERRAVQEM 2691

```

```

Qy 2057 AKE-----LSEIMHEQ 2067
Db 2692 SMNNDPTQSETHSE 2706

RESULT 9
HMW2 MYCPN
ID HMW2 MYCPN STANDARD; PRT; 1818 AA.
AC P75471;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome high molecular weight protein 2 (Cytochrome accessory
protein 2).
GN HMW2 OR MP310 OR MP526.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxId=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfeich R., Hilbert H., Plagens H., Pirk E., Li B.-C.,
RT Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
pneumoniae."
RL Nucleic Acids Res. 24:4420-4449 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97252497; PubMed=9098066;
RA Krause D.C., Proft T., Hedreya C.T., Hilbert H., Plagens H.,
RT Hermann R.;
RT "Transposon mutagenesis reinforces the correlation between Mycoplasma
pneumoniae cytoskeletal protein HMW2 and cytochrome."
RL J. Bacteriol. 179:2668-2677 (1997).
CC -I- FUNCTION: COMPONENT OF THE CYTOSKELETON-LIKE STRUCTURE WHICH
STABILIZES THE SHAPE OF THE WALL-LESS MYCOPLASMA. THIS
CYTOSKELETON-LIKE NETWORK OF ACCESSORY PROTEINS CONTAINING HMW
PROTEINS 1 TO 5 ALLOWS THE PROPER ANCHORING OF CYTADHESIN PROTEINS
IN THE MYCOPLASMA MEMBRANE AT THE ATTACHMENT ORGANELLE (BY
SIMILARITY).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib.ch).
CC -----
DR EMBL; AB000051; AAB96174.1; -.
DR EMBL; U59896; AAB52527.1; -.
DR PhosSite; P75471; -.
KW Cytochrome; Structural protein; Coiled coil; Complete proteome.
FT DOMAIN 31 880 COILED COIL (POTENTIAL).
FT 919 1607 COILED COIL (POTENTIAL).
FT DOMAIN 1644 1755 COILED COIL (POTENTIAL).
FT 1786 1817 COILED COIL (POTENTIAL).
FT DOMAIN 1818 21562 MW; 66DP4B08F0CPBC0 CRC64;
SQ SEQUENCE

Query Match 1.7%; Score 186; DB 1; Length 1818;
Best Local Similarity 17.5%; Pred. No. 0.017; Indels 710; Gaps 88;
Matches 369; Conservative 310; Mismatches 716;

```

QY 312 DSYLPEIASKARAEAIKLKSESSEVUKLFYLDPAQKLDFFSSABPEVKSFEKEFKRILVVC 371
 Db 196 DALVEYDHLVNL-----YELENOKRLVIGIEVATYQDVLVSADAELOVNETTIO----- 246
 QY 372 NDSFNLQCCVAENEBEPTTNVEPFVYTLSPDIKYNRKISADPHVDLNFHSYRQMIAT 431
 Db 247 NQANFQKQC-----DAYVAOLK-----QVEOOIQTT 272
 QY 432 SPALMNGSGPETOSALRGILHEAAMQPKQIGFVTCPPHPDIFLVARIEKVLGSGITHCA 491
 Db 273 KQSLV-----DEESTLKVRLNDA-----DFYINSRLAE-LDDLTSKIN 309
 QY 492 EPMKSSDSSKVAOKVLKAKQAQORIGQYRMPFAMAARTLFFDASGNLDKNARFSAIYR 551
 Db 310 ERFVFSKQADYKASLANLTKEKERLSAKQSFEBRLNTALDINRMEQENALFKAHLE 369
 QY 552 Q---DSNKLSDNMLKILADFR-----KPEKRAKLFVIL-----GNLDIT 589
 Db 370 QOQYEFERKQOESILKLETHEKLOKRIGEFKIESEKSEALLIORRELEKRRERIDDL 429
 QY 590 DNVSJD-----FPRYVNS-----YIPTKQ-----FENC 613
 Db 430 TQMSLEYEQORRNQVULKEKGRVOQHFOVLVHAKKLDQKRYHLABQKIDEEQIFKL- 488
 QY 614 SKTPIFEVEEFPVPCIPKHTQPYTIYNHLYVPKYLK-----YDSOK 656
 Db 489 -KEKITERRELEKVLVVKQKQDQKENDLLIFEKOLROYQADFENBIEKQONELFASQK 547
 QY 657 SPKAKNIALCIFKQSDSEDSOPKCIYGRPGGVPFTTSAPAAV---LHHNQPFYDE 713
 Db 548 SLOKS-----FTLOKKEALNOKAKI-----AEDMAHLKQNKHHADEIF-- 590
 QY 714 IKLELFTOLHEKHLTLTFPHVSCDNSKSTKRDVETQVGYMLPLL-----KQOR 767
 Db 591 LBEFPHNLOEKKLL-----EARTQFQNRVS-----LISAFKQKQAE 629
 QY 768 VTTSEOHIPVSAVLPGSYLGQYELGMRHYGPBIRKAVDGGKPLIKISTHVLSTVYTOQH 827
 Db 630 LVKQKQSL---EQLTAFNKEGE-----AVERDKDRLANLEKQKEMLGDKVHQFDEN 679
 QY 828 LNHFPQYQCTEGAGALGNELVKYLSLHAMEGHVMIAPLITLNOFRVLTQAEVY 887
 Db 680 SLNISKKLARELAIFKEKELEMAQKQSLDNNN--NAGLKQLDLKLSBSL-KTERLEI 736
 QY 888 AVNVTREVI-----IHVVAQCHEEGESHLRSYVKTAYVAEPYAS---EYKTVHELT 937
 Db 737 EASKEHILDFYDESSRIAD-YESDLOARL-AEVKLEKQOETAKSERELKVALEKLN 794
 QY 938 KSMTILKPSADELTJNKLUKTSWFFDVLIKSMACHLINSVUKLLRNQRPASYYHAAV 997
 Db 795 QAKKAFILO-----IRKQOLLE-----IASVQKQLOQ--KANLLKQOQ--AELDQOT 836
 QY 998 EYVVMNLMRHITQKFPDNPEASKNANHSLAVFIKRCFTFMQDGFVKQIINNYSQAPGD 1057
 Db 837 BELEAAFLRODDTK---KELEKALHS--VSKQOELBERESFLLQKOREFA----- 882
 QY 1058 PKTLFYEKFEFLVVCNHEHYIPLNLPMPGK---GRIOR-----QODLOLDVSLND 1106
 Db 883 -----EHVAGFKQV--HFKTTQMORLSEFNKQOQSEQIKRETELKIAFADLKQDOLP- 934
 QY 1107 EFCRNHFLVGLLREVGTALQEFREVRLIAISVLKNLLIKHSFDRVYASRSHQARIATLY 1166
 Db 935 ELQKN-----QEFQOIQKHKLELLA--QKQABLKQBLEQKATLALAG----- 976
 QY 1167 LPLFGLLIENVQRIINRVDSPPVNAGMTVYKDESLLAPVNPVLVTQKSGTLDNSLHKDL 1226
 Db 977 -----DQD-----TVOAKLIDLARQOHE-- 993
 QY 1227 LGALISGIASPYTTSTENINSVRNADSGSLISTDSGNSLPERNSEKSNLSLDKQOSTLG 1286
 Db 994 -----LELRQNAFQASLSLKKQREQLTNQVKVLHGEKGRHEKTLTK 1036

QY 1287 NSNVSCDKLDQSEIKSLMLCFLYIILKSMDDALFTYNNKASTSELMDEFTTSEVCLHQFO 1346
 Db 1037 DDLAEKEKQDQK-----KQABINQRFQFENEYAD-----FD 1069
 QY 1347 YMGKIYARTGMHARLQOLGSLDNLSTFNHSGSDADVLHQSLLEANIATEVCLTALD 1406
 Db 1070 QAKKR-----ELOELNQRIRNLB-----OSNMSLKK----- 1096
 QY 1407 TSLFTLAFKQNLADHGNPMLKKYFPDYVLCFLQHQSETLKAVFTALRSLIYKFPST 1466
 Db 1097 -----RNQLTIDFA---LKRKY-----QHTQTRVQVLNTQIKFLELEKN- 1134
 QY 1467 FYEGADMCQ---ALCYELIKCNSKLSIRTE-ASOLLFPLMNNFDYTGKSEFVRTHL 1522
 Db 1135 -FOKSDBAALQKALLIKRLRSGFASKLQORBALAQKLEFQKR---DQOQSEINNAKL 1190
 QY 1523 QYIISVQLIADVVGIGTRFQOQSIINNCANSQ---RLIHGTFSSDVKDLTKRIRV 1579
 Db 1191 Q---LEOFKLEKQNFDEAKQKQLEFQDQCORLQVERKRLK----- 1228
 QY 1580 LMATAQMKHENDPEMLVDLOYSLAKSYASTELAKTMLDSMARLHVKNQGLSEAMCYV 1639
 Db 1229 -----OKLVOLK-NLSKSYLT-----YKRRADLSQOOLQHK 1258
 QY 1640 HVTALVAEYLTRKEAVQWEPPLLPSSHSAQLRSRGV--FRQGTAFRVTIPNIDEAS 1697
 Db 1259 Y-----ANLBELKEKLTQAKRALDKKRAIYQKMAQFSELRQEKQKOLLSAQKQVDDKR 1313
 QY 1698 YMEDVGMQDVHNEVDVLELLEQCADGLMKARVYLADIYTLIPIYEKRRDPERLNL 1757
 Db 1314 LLEQ-----NORHLONLSET-----KKRGSLHDHINK---FQRRK----- 1348
 QY 1758 YDTHRAVSKYTEVHNSGRLIGTYFRVAFQGAQOYQFTDEBTVEGF-----F 1807
 Db 1349 -----EAVSSIANSKKL-----KQEGELQIGLOKLSLKQTOI 1382
 QY 1808 EDEDEKEYITYKRP---KLTPLSEISQRLLYSDKFGSENVMIQDSKVNPKDLDSKYA 1864
 Db 1383 EQFESKLYQQRREKLDRORTTLISKL-HRELKAQNEBATAHKNREVLB----- 1426
 QY 1865 YIQVTHVLPFPEBEKLOEKRTPEFSNINIRMPFE-----MPTQOTOK 1907
 Db 1427 -----LENYKKLOLRTTEKSEFDDNNKRLFEYFRKIRNEIEKKAHITVLEETOK 1479
 QY 1908 ROGVEEBOC---KRRITLFIHCFPYVKRIRPWYQH---HTDLNPI----- 1948
 Db 1480 KHLVETEVLKHLQKOSTISKGQELKEIKEVSRDISHTKQRELSLSLHQKLLQKN 1539
 QY 1949 ---EVAIDE---MSKVAELRQLCSSAEVDMIKLOLQOQ-----SVS 1985
 Db 1540 LAEREREINNKKSLLTQIKOTAKQKLSKEKARILKLEKMRVQEQOYQAEITRLKTRNAD 1599
 QY 1986 VQVNAQPLAVARAFLDDNTKRYP-----DNKVLLKEVFRQVFE----- 2025
 Db 1600 LEKNKNKHLFPLFLKINGNDMNYPPYPWPYPOQKQEDSSNQIRHLFQOQLQPMOQRYEN 1659
 QY 2026 ---ACGOALAVNERLIKEDOLEYOEMKAN---YREMAKELSEIMHBOICPLEKTSYVL 2078
 Db 1660 ELTELRRQALILEKLDQIOLESQSLAKNDPEKVEQMMQKLLTEQKLSAFQKINAL 1719
 QY 2079 PMSLH 2083
 Db 1720 AEOIN 1724
 RESULT 10
 CENE HUMAN
 ID CENE HUMAN STANDARD; PRT; 2663 AA.
 AC 002324;
 DT 01-JUL-1993 (Rel. 26. Created)
 DT 01-JUL-1993 (Rel. 26. Last sequence update)
 DT 30-MAY-2000 (Rel. 39. Last annotation update)
 DE Centromeric protein E (CENP-E protein).

GN CENPE.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RN NM_001080691.1
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024922; PubMed=1406971;
 RA Yen T.-J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
 RT "CENP-E is a putative kinetochore motor that accumulates just before
 RT mitosis.";
 RL Nature 359:536-539 (1992).
 RN NM_001080691.1
 RP CHARACTERIZATION.
 RX MEDLINE=95196755; PubMed=7889940;
 RA Thowar D.A., Jordan M.A., Schaar B.T., Yen T.-J., Wilson L.;
 RT "Mitotic Hela cells contain a CENP-E-associated minus end-directed
 RT microtubule motor.";
 RL EMBO J. 14:918-926 (1995).
 [3]
 RN RP
 RP CHARACTERIZATION.
 RX MEDLINE=98437347; PubMed=9763420;
 RA Chan G.K.T., Schaar B.T., Yen T.-J.;
 RT "Characterization of the kinetochore binding domain of CENP-E reveals
 RT interactions with the kinetochore proteins CENP-F and HUBB1.";
 RL Cell Biol. 143:49-63 (1998).
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
 CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
 CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
 CC AND/OR SPINDLE ELONGATION.
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBB1 KINASE.
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
 CC CONGRESSION. RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
 CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.ebi.ac.uk/announcements> or send an email to license@ebi.ac.uk).
 CC
 CC EMBL; Z15005; CAAT8727.1; -.
 DR PIR; S28261; S28261.
 DR HSSP; P17119; 3KAR.
 DR Genew; HGNC:1856; CENPE.
 DR MIM; 117143; -.
 DR InterPro; IPR001752; kinesin_motor.
 DR Pfam; PF00225; kinesin; 1.
 DR PRINTS; PRO00380; KINESINHEAVY.
 DR SMART; SM00129; KISC; 1.
 DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
 DR PROSITE; PS00417; KINESIN_MOTOR_DOMAIN2; 1.
 KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
 KW Cell cycle; Centromere.
 FT DOMAIN 1 335 KINESIN_MOTOR.
 FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
 FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
 FT NP_BIND 86 93 ATP (BY SIMILARITY).
 FT NP_BIND 2663 AA; 312087 MW; CEFCL3880C8CB8 CRC64;
 QO SEQUENCE

```

Query Match      1.7%; Score 186; DB 1; Length 2663;
Basic Local Similarity 17.1%; Pred. No. 0.031;
Matches 441; Conservative 380; Mismatches 949; Indels 814; Gaps 112;

QY   19 ALSKPGTAAELRQS--VSEVVRGSLAKPLIELD---YENVIVQKKQIINDCLRE 72
      |||:::||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db    319 ALGFASIKWKMKPPYVNEVSIDALLRKRYK-EIMDLXKQLKEESLETRQAQMEKQLA 377
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

QY   73 MLLEPYD---DFQNILRRQGRYCSTYPALAAEEBAQS--LFTPECIKTYNSDMHLYNY 126
      |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

```

Db	378	QLBEKOLLQVQNEKJENLTRMLVTSSLSLLOELAKRKRRTVCLGKJN-----XM	431
Oy	127	KYEDYSEFPRLPNKVVVKLDKLPVHVY-EVDEVDKDEDAASLGOSKJITKGV-----	180
Db	432	KNSVADQOFNIPNTITTKTKHLSINLREIDESCSESVFS--NTLDTLSIEMNPATK	489
Oy	181	-LYGNNNSAISVMTBSFKRFFHLLQLGDSYKFEF-LND-----LOKPKGSIF	229
Db	490	LJNONENTESELNSTRADYDNLVLDYELRTKEKEMELKEKNDLDEPFEALERTKDOE	549
Oy	230	LGFLYGVSPFRNNKRRFAFELKMODKSYLLAABSEVEM-----BEWTLINKIL--OLNF	283
Db	550	MOLIHESLNKLNVKH--REYVNDLLENL-----SSKELLREKDDQKKQOYIDSOKE	604
Oy	284	EAAMOEKRNQDSHDEOSK---LESGSGSLDYLPELAKSAREBAIKLK-----	330
Db	605	NIKDDLSYESIEDPKOMKOTLPDAETVALDAKRESAFLRSENLLEKMKKELATTVKQ	664
Oy	331	SESRYKFLYDPRQKXDPSAEBVVS--PEE--KPKRLLVK--NLSFLYLOCCVAENE	386
Db	665	MENDIOYQSOLEKAKKQVDELEKLOSAFNEITKLTSLIDKVPKOLLCLLE-----	718
Oy	387	EGPTTNEPFFVTLSPDKIKYNRKISADPHVDLHNFSPOMIATSPALMNGSPETQSA	446
Db	719	EGKTTDLOK-----BLNKVE-----ENELREVVILLSEL--KSLPSEVR	758
Oy	447	LRGILHEAMQYPKQGISVYTCPPHPIFL--VARIEKLYOGSITHCAPRYMSDSSKVAQ	505
Db	759	LR-----KEIDKSEELHITSEBKDKFSEVHVHESVQGL-----BEIGTKDRLATQ	809
Oy	506	KVLKNAQACORLQCYMPRMAAARTLFKDASG-----NLDKRA-RPSAIV-----	550
Db	810	SNVYSTQOEPFNFKTLHMDFEQKYKMYLEBEMRMOEIVNLSKEAKQFDSISGALKTELS	869
Oy	551	-----RODSNKLSNDMLGLADPRK-----PEMKK-----	577
Db	870	YKTOELOEKTRBEVOERLANMEQLEKQOENRDSPLQTYBERKTLITELKOOLBEEVKTLQ	929
Oy	578	-----LPVILGLDITIDNVSSDPFNYSVSIYTPKOF-----ETCS--KTPITP	620
Db	930	EKDDLKQLOESLOERQDKSIDHDYVNMNIDQEOURLNLESJKOIHETINTLKSRISE	989
Oy	621	EVEEFPVPCIPKHOTPYITYNHLYVYRYKLYDSQSKPAKARNIATCIEFQDSBEDSOP	680
Db	990	EVSNNL-----HMEENTGETKDEF--QQKMGVIGDKDOLLEAKNTOTLPADVDKNEIEQO-	1042
Oy	681	LKCIYGRPGGVFRSRAFAVLHNNQRPEDEL--KIELPQOL-----	722
Db	1043	-----RKIFSLIQEKNELOQMLESYIAKEOOLKTDLKNENIENTIENOBELR	1088
Oy	723	-----HEKGHLLLTFFHYS--CDNSKSKSTK-----KR	748
Db	1089	ULGDELKQOEIYAQEOENHAIKKEGELSRTCRLAEBVKLEKESQLOEKQOOLLNVQE	1148
Oy	749	DVETQVGYSMPLPKD--GRVVTSE---QHIVSANLPSSGYLYQOELMGHHYGEI	801
Db	1149	EMSEMOKKINEIKELKMKELTELHMETETERLELAQKLNENY-----BEV	1195
Oy	802	KMVDGGRPLKISTHNVSTVYTOQHILNFRYOQKTESGQALGNELVTKLSLHMEG	861
Db	1196	KSITKEKRVLK--ELQKSPETERDHLRGYIREIAT-----GLQTKBELKI--A	1240
Oy	862	HVMIAFLPTLINOFLRVILTRAT-----QEVAVNVTRVILHVAOCHEE	905
Db	1241	HIHKLKEHETIDELARSVSEKTAQINTODLEKSHNTQOEBEIP-----LHE-	1287
Oy	906	GLSEHLSYVKAYAKBAPYVASEKYVHEELJTSMTILKPSADFLTSNKLKYSWFFD	965
Db	1288	--BOELLPNVKVASETOETMNEBELLTGEOSTTKOSTLIARIMERLRYLKEFQSS---Q	1341
Oy	966	VLIYSMAQHLIENSKVULKNQRPASYNHNAVEIVNNMLNPHITQKRDNPBEASKNANHS	10255
Db	1342	BEIKSLRK-----ERDNLTIK-----BLEVYKQDLKEHIRETLAKIQESQKQEOQS	1389

Qy	1026	LAVFIRKCFTEPMDFVFFVFOJINNNYSICFAPGPKTLFEFKPEFL----	RVCNMHHYIPL	1081
Db	1330	L-----NMKEKNETTKIVSEMEOKPKD-SALLAIIEIMLQSKRLOESHDME--	1433	
Qy	1082	NLPMBFGKRIQRYODLOLDVSLTDFCNMHELVGLIREVGTALQEFREYALIAISVK	1141	
Db	1438	-KSAVEKODIQRLOEVL--QESDQLEN-----IKETIAKLETEEBELKVHCHCK	1487	
Qy	1142	NLLIGHSDDDRASNSH-QARATIYLPFL-----LIENQIRINRVDSPPV	1190	
Db	1468	EO--EETINELRVNISEKETELISTQKOLEALNDKLOKNOIEIYEKEQNLINQISEVOE	1545	
Qy	1191	NAG-----MTVD-----ESTALPAVNPVLTPOK-----	1215	
Db	1546	NVNEIKQEKERHKAQDASALQSIESTKMLELTPLOESOEIQIMKEKEMKRVGALQIE	1605	
Qy	1216	-----STLDSLHKDL-----LGAISGIASPYTSTPNINSVRN	1249	
Db	1606	RDQLKENTKEIVAKKKEQSEKEQYGLKMTAVIETQEKKEIEHLKEQETQKLNENIET	1665	
Qy	1230	ADSRSLSTDSGNSLPE-----RSEKSNLSLDKQOQSTLGNISVVRCDKLDQSEIK	1301	
Db	1666	ENIRLQIOLHENTLEEMRSVTKERDDLRASVEELTKVTERQOLKEMNLETTIR--DLEKQEBLK	1724	
Qy	1302	SLMCFEYLKMSMSDALFTYNNKASTSELMDFITISECJLHQFYQMGKVIARTGMNHA	1361	
Db	1725	IYHM---HLKEHOE-----TIDKL-----RGIVSE	1746	
Qy	1362	RLQOQGSLDNSLTFFNHSYGHSDADVLHOSLSEANITATEVCITLALDLSLFTLAFKNOJLLA	1421	
Db	1747	KTNEISNMQKDE-----HSH-----DALKAQDLKIQEBELRI	1778	
Qy	1422	DHGHPLPKKVPDYVLCFLQKQOSTALKANFTAIRSLIYFPPSFTFYGRADMCAALCYE	1481	
Db	1779	AHMH-----LKEQOETIDK-----LRGIVSEKTDKLSNMQXD-----	1810	
Qy	1482	ILKCCNSKL-----SIRTEASQLAFLFRNNDFYGGKSFVRTHLOVLIISQOLADVVG	1537	
Db	1811	-LENNAKIQEKIOELKANEHOLI--TLKQDNQETQKVSSEMOUKQDQIKOOSTLSCLE	1867	
Qy	1538	IGTRFQOGL-----SIINNCANSDDL-----IKTSPSSDVLDL--TKRIR	1577	
Db	1868	IENLNLQOELHENTLEEMKSVMKERNLNRVEETTLKLBRODLKESLQETRKADLDLQOGLK	1927	
Qy	1578	TVLMTAQMKHEANDPEML-----VDLOYSLAKS-----YASPELRKTYLDSMA	1622	
Db	1928	TARMLS-----KEKETVTDKLRKISEKTIQISPIQDLDKSKDQLOKKIOELQK--KELQ	1981	
Qy	1623	RIHVNGDLSSEAKUCYVHTALVAEYLTRKEAVQWEPRLPHSHSACLRBSGCVFRQGC	1682	
Db	1982	LIRLVK---EDVNMHKKINEM--EQLKK---QEPNYLCKCEMDFOLTK--KLHESL	2029	
Qy	1663	TAFRVITENIDEASWMEVDGMQDVHF-----NEDV-----LME	1716	
Db	2030	EHIRVAKERDELRIKESLSKMERQFATLRBMAIRQRNHQVPERKRLSDQGOHME	2089	
Qy	1717	-LLEGCADGLMKAKEYELADIYKLIPII--YEKGRPERLAHLVDTLHRAYSKTEVM	1772	
Db	2090	SLREKSCRIKELLKRYSEMDHYECLNLSIDLEKEIEFHRIMKKLKYLSVYTIKIEBO	2149	
Qy	1773	HS-----GRLL-----	1778	
Db	2150	HECINKFEMDFDEVKQKELLIKIQHLQOQCDVPSRELRLDKLNQNMDLHIEELIKDFS	2209	
Qy	1779	-----LGTYFRVAFFGQAOYOFTD-----SETD-----VEGEF--	1807	
Db	2210	ESEPSITTEHQVULSNKEMTQFLBEWLNTRFDIEKJKNQIKQENDKICQVNNFFNRI	2263	
Qy	1808	-----EDDEGKEYIYK--EPKLTPLSEISQRLKLKYSDFSGSENVKMTIQDSG-KVNP	1856	
Db	2270	IANNMESTEFERSATISKEMWODLSKLEKKEKEXLFKQY-----QTLKTSILASGAQVNP	2323	

QY	1657	KDLDSKVAAYID-----VTHNIPFEDKEKLEOERTEFEREHNIRFMFEMPFTO--TGKR	1908
Db	2324	TTQGNKMPHYVSRATQLTTEKIRLELNSLHEAKESANHKSIIKXQKELEVNDDIATL	2383
QY	1909	QGVVEQCKERTILTALHCPFPYKGRIPMYQHHTDNLPIEVALIDEMSKVYAEIROLCSS	1968
Db	2384	QAKHESNK-----CLEKTEKTIQVLOD-----KVALGAKPYK-----	2416
QY	1969	AEVDMIKLQTLQGSVSGVWAGPLAVARAFLDPTN-----TKRPDMKVLLKEVPPOFV	2024
Db	2417	FEIEDLNMKL-----VKIDLEMKNAKEPEKISTAKTAVEYQKQVILLRENLRRSQ	2469
QY	2025	EA-----CGQALAV--NERLLIKEDOLEYQEMKANRYEMAKELS	2061
Db	2470	QAQPTSVISHETDPOPSNKPLFCGGSGGIQVONTKALI-----LKSEHIRLEKIS	2519
QY	2062	EIMHEQICPLEEKSVLPNSLHIFNAL-----SGTPISTWYHGHTSS	2103
Db	2520	KLKQONQOLIKQKELLNNQHSNEVKTWKERTLKRAHKQVTCENSPKSPKVTGASK	2579
QY	2104	SSVV	2107
Db	2580	KKQI	2583
RESULT 11			
ID	DVAL	DICVI	STANDARD; PRT; 1557 AA.
AC	Q24702;		
DI	01-NOV-1997	(Rel. 35, Created)	
DT	01-NOV-1997	(Rel. 35, Last sequence update)	
DT	01-NOV-1997	(Rel. 35, Last annotation update)	
DE	DVA-1	polyprotein precursor (Antigen-1) (Allergen-1) (NPA).	
GN	DVA-1.		
OS	Dictyocaulus viviparus (Bovine lungworm).		
OC	Eukaryota; Metazoa; Nematoda; Chromodorea; Rhabditida; Strongylida;		
OC	Trichostrongyloidea; Dictyocaulidae; Dictyocaulinae; Dictyocaulus.		
OX	NCBI_Taxid=29172;		
RM	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=96089829; PubMed=8538702;		
RA	Britton C., Moore J., Gillear J.S., Kennedy M.W.;		
RT	"Extensive diversity in repeat unit sequences of the cDNA encoding		
RT	the polyprotein antigen/allergen from the bovine lungworm		
RT	Dictyocaulus viviparus.";		
RL	MoJ. Biochem. Parasitol. 72:77-88(1995).		
RN	[2]		
RP	FATTY ACID/RETINOID-BINDING.		
RX	MEDLINE=95370256; PubMed=7642601;		
RA	Kennedy M.W., Britton C., Price N.C., Kelly S.M., Cooper A.;		
RT	"The DVA-1 polyprotein of the parasitic nematode Dictyocaulus		
RT	viviparus. A small helix-rich lipid-binding protein.";		
RL	J. Biol. Chem. 270:19277-19281(1995).		
CC	-1- FUNCTION: HAS HIGH BINDING AFFINITY FOR FATTY ACIDS AND RETINOIDS.		
CC	-1- SIMILARITY: TO A.SUM ABA-1.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; U02568; AAC47403.1; "		
KW	Signal; Polypeptide; Lipid-binding; Retinol-binding; Repeat;		
KW	Glycoprotein; Antigen; Allergen.		
FT	SIGNAL	1	21
FT	CHAIN	22	1557
FT	PEPTIDE	61	213
FT	PEPTIDE	214	334
FT	PEPTIDE	335	480
FT	PEPTIDE	481	606


```

FT PEPTIDE 607 737 E (POTENTIAL).
FT PEPTIDE 738 869 F (POTENTIAL).
FT PEPTIDE 870 1003 G (POTENTIAL).
FT PEPTIDE 1004 1074 H (POTENTIAL).
FT PEPTIDE 1075 1208 I (POTENTIAL).
FT PEPTIDE 1209 1342 J (POTENTIAL).
FT PEPTIDE 1343 1475 K (POTENTIAL).
FT PEPTIDE 1476 1545 L (POTENTIAL).
FT PEPTIDE 1546 1557 C-TERMINAL EXTENSION (POTENTIAL).
FT CARBOHYD 997 997 N-LINKED (GLUCNAC. .) (POTENTIAL).
FT VARIANT 946 946 H -> Y.
FT VARIANT 948 948 L -> V.
FT VARIANT 955 955 H -> D.
FT VARIANT 974 974 E -> G.
FT VARIANT 990 990 G -> A.
FT VARIANT 1543 1543 A -> V.
SQ SEQUENCE 1557 AA; 182600 MW; 33717911CF55BDAB CRC64;

Query Match 1.7%; Score 185; DB 1; Length 1557;
Best Local Similarity 16.2%; Pred. No. 0.015;
Matches 269; Conservative 262; Mismatches 582; Indels 546; Gaps 69;

QY 490 CAEPYKSSDSSKVAO-----KULNAKQACQRLGQYRMPAMARTL-----FKDA 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 214 CKKYFMQJANSDEVEKIKSLNDEIRHVYKNAVARLNGELKEFAVVMETLCEVDLAFKAR 273
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 537 SGNLDNAFSAIYRQDSNKLSDMDLKLADPR-KREKVAKLPIVLGDLITIDVSSD 595
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 274 KNDIDDKINRLISMWTEOK---QVVKQYADGRSADIFAKIPEFLSIDGAGYAAR- 329
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 596 PNNVNSSYIPTKOPETCSKPTPE-----VEEFVPCIPKHTOPTYITNHL 643
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 330 -AQIQECCYKMBEVATAEIALHEIHEDHDCGRKVFETIGRLPE----- 376
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 644 YVYPTKLKDSQSKPAKANNAIC--TEKSDDEBSQPKICYGRGGPVTRSAFAAV 701
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 377 -----DRKLEVEKDLPCCEKIMWYRDHGDHNSHK-----HGA 407
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 702 LHHNHN-----PEFYDEIKELPTOLHEGHNLILPFHVSCNSSGSKTKKD 749
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 408 HHHHHHLAVRRRHLYAIKFLDMLK--PEQGHLEKI-----SNS--GAFFDV 453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 750 VVETOVGYSMLPRLKDGAVVTSEONIPVSANLPSGYLGYOELGMGRHYPEIKWVGKRP 809
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 454 IAEVKKFYGLP-----BEKIELKAKFKSQCYDW----- 483
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 810 LKISTHLVSTVYTTODQHLNFFQYQCKTESGAQALGNEIVKYLKSLHAMEGHVMIATLP 869
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 484 -----VKEVATSEEM-----NDIMK-----MHESKNH-----S 506
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 870 TILNQLFRLVLRATQSEVAVNVTIRVIHYVAOC-----HEEGLSHLSRYVYK 917
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 507 DLMKLTLENLTEDEQ---KHTLEHREVCGLMEVQNTNKKQKQSLSEAMDYLSW 561
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 918 AYKAEF-----YVASEYKTVHELTLSMTTLKPSADELTSNKLKYSMPFEDVLKSM 971
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 562 MTEDEKEXKAIYETSNRQTFPYDELK-----IMESSEDEVAKATK-----LEAAKRY 612
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 972 AOHLIENSVKLLRNORFPASVYHAAVEYVNMIMPHITOKFRDNPASQANSHSLAVFIK 1031
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 613 GTNIIGEEVNDIIRKMKKNGATPEEISNRVDELIGETSDR-----KEKAYRMSKICK 666
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1032 RCFETMDRGFVFKQINNYISCFAPGDKTLFEKPE-----FLRVVCHN 1075
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 667 KIYSL-----GHSKQLOQYDFENVLQKYLTLWLDQSKNELRTMSDN 707
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1076 EHYIPLNLMPPF-----GKGRIORVQLOLDYSLTDFECRNHFLVGLLREVGTALOEPR 1131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 708 KKKIKYKIIDYFDGTIGEVKEKAVELOL-----ACNHYIKSLVGEKAMEIKQLKE 759
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1132 VRLIAISVLKNT--LIKSPDDRYSRSHQARIATLYLFLGILLNNVORIN----- 1181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 760 BCKSSSEIKKVEDVINQISDSIRSRADALL--VCKRIFG-IYRLRLRDNSEIHSLEE 816
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY 1182 -----VRDVSPEPVNAGMTVKDESLAPAVNPLVTPQKSGT-----LDNSLH-- 1223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 817 AMERLYTLMSDDQKIVISYDVNDRKLYEKLMEFPFDALGETQOXAKEIKDACKIV 876
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1224 KDLGAIISGLASPYTTSTPNINSVYVADSRGLISTDSGNSLPERNSKNSLDRHQSS 1283
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 877 KOLIGENG-----NLREMKENGA--SNEAIATVEEMIAITDETGRAQM 922
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1284 TLGNS-----VVRCDKLDQSEIKSLMCFYLILKMSDDLL--FTWNKASTSELDNF 1334
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 923 RASTSCRRVYGVQFRDRDHHEHNL-----DEALEKFTWLNBEQSKQLK-- 968
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1335 FTISEVCLHQFQWYMKRYIARTGMHARLQOLGSLDNLSTFNHSYSGSDADVLHQSLEA 1394
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 969 -TI-----YESDEBALHKKWMEF 986
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1395 NIAVEVCLTALDTLSLFTLAFKNOQLADHGNHPLKQVFDVYLCEFLQHQSETALKVNF- 1453
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 987 FEAGAGLRASVNSAKKIYGVV--KRPRDHHEHNLDEALEKYLTLWLNBEQ--KSQKITYE 1043
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1454 TMLRSLIYKFBSTYE---GRADMCALCYELKCSNKSLSIRTEASQLYLFLRRNPD 1510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1044 SGDRBALYKQVLEFEAATGEVKEKAAV--ELKSACRHYIK-----D 1083
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1511 YTG--KKSFVRTHLQVLIISQDLADVVGIGTRFQGSLSIINNCANSDRILKHTSPSSD 1568
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1084 YIGDEKAEKIKEMESGVSTEBI-----SKYDEFLAMITDBKAKALRASAACK 1135
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1569 VKDLTKIRITVLMATAQMKENHEDEMLVDLOYSLAKSVASTPELRKWTWDSMARHYK- 1627
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1136 IYGVAKRFR-----RDHHEH--NLEALEKYL-----TWLNBEQSQMKT 1174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1628 ---NGDLSEAMCYHNTALVAEYLTREKNAVOMERPLPHSHSACLESRGV----- 1677
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1175 IYESGDRBALYKQVLEFEAATGEVKEKAAVELK-----SACRHYIKDYIGDEKAEK 1226
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1678 ---FROGCTAFRVIIPNIDSEAMMEDYGMQVHNEDVLMELBOCADGLMKAREYELI 1734
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1227 IKEMESGVSTEBIKKYDEFLAMITD-----EKAKALRASAACKKI 1270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1735 ADIYKLIIPYKRRDFERLALYDLTLRAVSKYTEVNHSG--RLLYGTYFRVAFGQAAQ 1793
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1271 YGVAKRL-----RDHHEHNLLEAMGKYLSSWMSDEQAQYKKIYGTGDRLATYNNKWE 1324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1794 YQFTDSEFDVSGFDEDECKEYTYKPKLTPLEISQQLKLYSKPSSEN--VKMID 1850
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1325 -----LFESVPSDE--KEKATSQLKACRHYIK--DFIGKDLAVIKEMER 1366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1851 SGKVPK-----DLDSKYAYIO-----VTH 1870
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1367 SGATNEAIGEKIDEFLAGLDQKQAQOARASACKKIYGVKSRGRREHYEIDVDAISK 1426
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1871 VLPFDEKELOERK--TEFERSHNIRRMFEMFPYQTSKQGVGEQCKRRTILTAIHCF 1928
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1427 YLTWNEBQKAEIKQDKKDEKQOTIGCKKIMEFFELTSDDDEKAREBQLKAA-----CX 1479
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1929 PYVKRIRIPVW--OHHTDLNPIE--VAIDMSKTVAB 1961
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1480 HYVK-----MYGEBKAELKYLKDSGISLEMSKYTE 1513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 12
YFP73_CABEL
ID YFP73_CABEL STANDARD; PRT; 4385 AA.
AC Q09322;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Hypothetical 486.0 kDa protein B0228.3 in chromosome II.
GN B0228.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;

```


OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCB1 Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Briceol N2;
RA Leimbach D., Waterston R.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U23168; AAC3807.1; -
DR WormRep; B0228.3; CB01744.
KW Hypothetical protein.
SQ SEQUENCE 4385 AA; 486003 MW; 50B8871B6C45FA23 CRC64;

Query Match 1.7%; Score 184; DB 1; Length 4385;
Best Local Similarity 17.2%; Pred. No. 0.086;
Matches 436; Conservative 393; Mismatches 845; Indels 856; Gaps 118;

QY 11 AETRKFTRALSKPGTAELRQSVSEVRLAKPELLEPLDYENVIVQK----- 61
DB 174 ABRVRF-----AQLRASADEITIRVALGSMQLEQAFAFMSLLITSVRCGLRTI 223
QY 62 -----KQIQLNDCLREMLLPYDPOFRIILRQGR--YICSTVPAKKEEQLFVTE 112
DB 224 APSNIIANTEIFDYVAEKKMSV-----SGVMKRERKEHSSKKFTTSREELIIGFWKGE 277
QY 113 -----CIKTYNSDMHLYNRYKE-----DYSGEF-----RQLPNKV 143
DB 278 RDEBRVVKLKD--RNVSPFHSLSVEAASSTSENVSLGKRKADQKSEFTIISQLSKREIV 335
QY 144 -----KLDELPHYVYEVEDEVDKDEDAISGQKIGTHGWLKGMNMSAISVTM 194
DB 336 SEAYGSESKLDQF-----FQVMEKMDWSN--TELGEKE-----HTALSANI 375
QY 195 RSKRRKRFPHILQGDGSKYKFEPLDKQEKKSIFLGFLYGVSPFNKKRRRFAFEIK--- 251
DB 376 RT-----LAPROKICDG--ILGKLKAPKQODESVGTQIQEIRRAI 413
QY 252 --MODSSYLLADSEVEMEWITILNKILQLNFEAMQEKRGSHDEQSKLEGSGS 309
DB 414 AVMSVPAASLLITSSNSNFSKMSISEKAVFSNLIVT-----TSHNLSITGTSSTTT 467
QY 310 GLDSY--LPELAKSAREAEIKLKSRSRYKLFLYDPDAQKLDFFSASPEVKSFEKFKRI 367
DB 468 ASISYODIPEMLAKS-----LWISRSEKLEKKEIRREPVIQTVESFW----- 509
QY 368 LKCNCLSPLOCCVANEGERPTTNVPPFVITSLDIDKNRKISADFHYDLNHSVRQM 427
DB 510 ---NTNDOEKVAVLNEK-----IDSIYSSLNTLAASWENELIQEILVRHTENAGKN 560
QY 428 IATTS-ALMNGSGPETOSALRGILHEAMQYKQGFISVTCPHPIPLVARIEKVLQGS 486
DB 561 IKRISRELVSSEPKTSSDL-----QQPFNV-----LEKNDMGO 595
QY 487 ITHGAPYKSSDSSKVAQKVLKNAQACORLQGYMPFAMARTLFKDASGNLDKNAEF 546
DB 596 I---SLPAQEHAIISKVRSLAPSASELSNIIGKLAPEMQEBETDLK-----INQAQ 646
QY 547 SAIRYDSKSLSDMDMLKLADPRKPEKAKLPVILGNDITIDVNSDPFNYSSTYIP 606
DB 647 ARVVLNVSSMSN-----TISSETFSRIP-----ENKATLTINY--GIIT 686
QY 607 TKQFET--CSKPTFEVEEFVPCIPKHOPYITYNHLVYVPKYLYKQSSQSAKANI 664
DB 687 TODLSLIGASSENQTIEMD-----YTEAKEELASAKLV--DKNLSVLKT--- 729

QY 665 AICIEFKDSDEEDSQPLKCIYGRPGGVFTSFAFAV-----LHHNQPEYDEIKIE 717
DB 730 ---EIRESGDEVQGF-----NWTASDEKEVGAIIICEKUSIHHTLHTAIRTVIES 778
QY 718 LPTQLHEKHILLTTFPHYSCNDSKSTKRPVVT-----QVGYEWL 760
DB 779 LSTDIQKEQKSLTHHFV-----KLST--RDVQAARISSEVDHLTLVEKMDWSKI 830
QY 761 PLAKD--GRVTSQOHIPVANSANLPFGYLYG-----QELGMGRHYGPEIKWVGSKPLK 812
DB 831 DLVEPVYSSISNVQALATSSQSSGIGLKLPPAPEETTSK---EIREINTAKILIN 886
QY 813 ISTHLVSTVYTDQDHL-----HNFQYQ 836
DB 887 VISSINSTI--TSDASLDMISEEQAIFSNIIGWASNNILTSSQGTSSFGFNVFELSE 945
QY 837 K-----TESGAQL-----GNELVKYLKSLHMGHWIAF--LPTIQLDFVLV-- 880
DB 946 ARVLAQESNRKQNLQKVRSESEELIHGIWSTASETEKVAIVKEKLETV--HQAMETLAIQ 1004
QY 881 RATOEEVAVN-----VTRVLIHVVAQCHEGLESHLRSYKYAYKAE 922
DB 1005 MATQ---SVNSSELISSEGNLASIKSTILPTREVISAGFISNEVQKULEVLTKEMWTI 1061
QY 923 PYVASEYKTVHEELTKSMITLIKPSADP-----LTSNKLKYSWFFPDVL 967
DB 1062 SLPOSEY---ESISQVYALAEFPNCDSLIGKINAPQOTETTSHLEIKNTAAVAVAN 1117
QY 968 IKSMAGHL--ENSKVLLRNRPASVHHAVETVNNLMRP-----ITQKPDNPA 1018
DB 1118 VAAAEVSYSKSSSLAKL-----PADEKALITNAGLVLSDVSLCTSSSTFPDORI 1171
QY 1019 SKNAHSLAVFIKRCFTFMDRGFVFKQJNNYISCF---APGDKTLFEYKFEFLRVGNH 1075
DB 1172 FENQANANISLQAVRETIILQR--LREPIENQVQGWSTASNOEKSAFLAK----- 1219
QY 1076 EHYIPLNIPMPFGKRIQRYQDLQDLSLDBEFCRNFPLVGLLREVGTLQAEFREVRIL 1135
DB 1220 -----QLDITMYD-----TMKVI 1232
QY 1136 AISVLKULLIKHSPPDRYASRSHQARIATLYPLRGLLIENVQRINVRDVSFPFVNAQGT 1195
DB 1233 AIQ-----MISQITIGDFSSASISSE-----TFKNDIKK--ARBV--VAAEFG 1271
QY 1196 VKDESL--ALPAPNAVLTVPQKSTLDNSLHKDLGALISGASPYTTSPNINSVNN--- 1249
DB 1272 IANESVQALAEVLNKV--EWSDISLPEKLHQNISSNAVVLAD---SGVNCDSIIIGKUNA 1325
QY 1250 -----ADSRGSLISTDSGNSLPERNSEKNSLUDKHQOOSTLGNS 1288
DB 1326 PEPQSAVYDQQTSEHQTELVIANIKSAVSVISNDSVLKRSKSE-----DEKAVISNIANVL 1380
QY 1289 VRCQDKLQOSSEIKSLMCELYLTKMSDDAL-----FT-----YWNKAS 1327
DB 1381 IISCDLASMSSTNSFEFORDILEPQANAVLGLTPASQVPTRNLOEPIQSEVQGWSTNS 1440
QY 1328 TSELMDFTTISEVCHQFOYMGKRYIARTGMM-----HARLQOLSLDMS 1372
DB 1441 SOE-----KASLVITQKLNKVKYDKAMKMAATEIAYQSINSINTISEENLENSQSPKDS 1492
QY 1373 L--TNNSHYGSDDADVLHQSLEANIATEVCITALDTSLEFLAKNQQLADHGN--PL 1428
DB 1493 IRESIMAFGVSSSEHV--QKTLLEISKEBELGITVPAVEYEAIVQNIKALSEPNCESII 1550
QY 1423 MKC-----VPDVIYLCFLQKHQSEFRLAKNVFAL-----RSLIYK 1462
DB 1551 LKGLNAPDPOGAWMI---INMEQRKLTIVNSIQAISVISNDSSEKLTLEVDSVLTK 1607
QY 1463 PFTSYEGRADICCALC--YEILKCNKSLSS---IRTEASQLLYFLNR-----NNPD 1510
DB 1608 ISEIIVSHDLTALSSCGDFNLQGSNTQEKADIDLTKTPNSGMLERAREPIEKQVQGW 1667
QY 1511 YTGKS-----FVRTHLQVITISVSQ-----LIADV-- 1536

```

Db      1668 STASNSEKQEMFLKEKVEITIHMLQTFSSALVSETVQRPDAVOSLALRSIOLTPREI 1727
Qy      1537 -----GIGTRFOQSLSIINN-----CANSDELIGHSF-SSDV-----KD 1571
Db      1728 LCAAGFISGEQLEQTFRGIDEDVMSHIDIPCALRDMLVNLAI VNVADANVNGNLISRPD 1787
Qy      1572 LTKRIRTVLMATAQMKENHEDPEMLVDLOYSLAKSASTPELRKTLWDSMARIVHNGDL 1631
Db      1788 ENQGETSTVL-----EEKNRIGIPLNLQRLSDQTYQMTSSLSLR-----SDEIEIVVNSMI 1836
Qy      1632 SEAMACYHVTALVAEYLTRKAVQMEPPLRLPHSHSACLRNRGCVFROGCTAFVITYN 1691
Db      1837 -----ISLISSENIL-----GDLISQ---AVOLAQPN 1859
Qy      1692 IDEASPMEDVGMQDVHPMEDVLMELLEGADGIMWKEVEYLADIYKLIPIYERGRPF 1751
Db      1860 MSDETEKSEFEL-PRNLLNLR-VVPEPSEBSIQSPFKTS---QLAEKASSTI-----S 1906
Qy      1752 ERLAHLYDLTHRAVSQVTEVMSHGRLLQTY----FRVAFQQAQVQFTDSETDV--EG 1805
Db      1907 EKLSTMLQ-----SEFVVSAAKQVSTSLTDVRRKIFPNQNSIITFGDLTRDVTYKA 1956
Qy      1806 FPEDEGKVIYKEPKLTPLEISQRLKLYSDKFG--SENVKMIQDS-----GKVPK 1857
Db      1957 FVSDEETLTLFSQLQSDMSOI-----KLSSKQKSLISANIKSLATSNLDTLIGQLMR 2011
Qy      1858 DLDSTAYVYQVTHVPIFPEKELQERKT-EFERSNIRIRPMPEPTQKQGGVEBOC 1916
Db      2012 EADSDQS-----EASLSERSVEMENTFKIQIITD-----TVVKKFSGDSE- 2053
Qy      1917 KKRITLTAIHCEPPYK-----KRIPWVYQHTDPIEVALIDEMSKVAELRQC 1966
Db      2054 -RASLINLIGLTSVNLKAIQVASEERLPDLSIDSTASTTPVFKQOENHROQTEGVQGF 2112
Qy      1967 SSAB-----VDMIKQLKQGSVSQVQVAGPLAVARFL-----DDNTKRYPD 2010
Db      2113 WSKHEHPNETVEVNVNRKIEVLKSI--LNCYVVAECHKLSLEKPSQODTECATRGVT 2169
Qy      2011 NVV-----KLKEVFR---QVEACQGLANVERLI-----KED 2041
Db      2170 EYVQNAFAVTEQSYSLKLVLDVSMONFELPEAVKETISKLKILNLPVDTLLVSKYKS 2229
Qy      2042 Q-----LEVOE--EMKANYR-----EMAKELSEIMHEQICPLEEKTSLVPSNLHPF-- 2085
Db      2230 HYSKIVEBELVQFMENFKSALNSEIENGQNSK-----C-IEETVTVLPNVALIASA 2282
Qy      2086 NAISGTPST 2095
Db      2283 NLATVPRES 2292

```

RESULT 13

```

A111_HUMAN
ID AK11_HUMAN STANDARD; PRT; 1901 AA.
AC 09UKA4; 075124; GN0K7;
DT 16-OCT-2001 (rel. 40, Created)
DT 16-OCT-2001 (rel. 40, Last sequence update)
DE A-kinase anchor protein 11 (protein kinase A anchoring protein 11)
DE (PRAV11) (A kinase anchor protein 220 kDa) (AKAP 220) (hAKAP220).
GN AKAP11 OR AKAP220 OR KIAA0629.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20323159; PubMed=10864471;
RA Reiton N., Collas P., Haugen T.B., Skahnegg B.S., Hansson V.,
RA Jahnson T., Tsekken K.,
RT "Localization of a novel human A-kinase-anchoring protein, hAKAP220,
RT during spermatogenesis."

```

```

RL Dev. Biol. 223:194-204 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Bates K.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1290-1901 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 5:169-176 (1998).
RN [4]
RP SEQUENCE OF 1121-1469 FROM N.A.
RC TISSUE=Placenta;
RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC - FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE A
CC AND ANCHORS/TARGETS THEM.
CC - SUBCELLULAR LOCATION: CYTOPLASMIC IN PREMEIOTIC PACHYTENE
CC SPERMATOCYTES AND IN THE CENTROSOME OF DEVELOPING POSTMEIOTIC GERM
CC CELLS, WHILE A MIDPIECE/CENTROSOME LOCALIZATION WAS FOUND IN
CC ELONGATING SPERMATOCYTES AND MATURE SPERM.
CC - TISSUE SPECIFICITY: EXPRESSED IN HEART, BRAIN, LUNG, LIVER,
CC KIDNEY, TESTIS AND OVARY. WEAKLY EXPRESSED IN SKELETAL MUSCLE,
CC PANCREAS AND SPLEEN.
CC - DOMAIN: RII-ALPHA BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC
CC HELIX, COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
CC COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC EMBL, AF176555; AAF07045.1; -
CC EMBL, AL136527; CAB89419.1; -
CC EMBL, AB014529; BAA31604.1; -
CC EMBL, AK002166; BAA92117.1; ALT_INIT.
CC DR Genew; HGNC:369; AKAP11.
CC MIM; 604696; -
CC FT DOMAIN 1650 1663 PKA-RII SUBUNIT BINDING DOMAIN.
CC FT DOMAIN 238 315 SER-RICH.
CC FT DOMAIN 1705 1767 SER-RICH.
CC FT CONFLICT 1303 1303 K->R (IN REF. 4).
CC FT CONFLICT 1444 1469 LDPYRNEVSQLYSFTSLVSHITKDA -> CGPSVELSPWK
CC FT MOTRCGRGNSQWKR (IN REF. 4).
CC FT MOTRCGRGNSQWKR (IN REF. 4).
CC SQ SEQUENCE 1901 AA; 210510 MW; 0893316C46A75672 CRC64;

```

Query Match 1.7%; Score 183.5; DB 1; Length 1901;

Best Local Similarity 18.7%; Pred. No. 0.026; Matches 337; Conservative 246; Mismatches 637; Indels 581; Gaps 84;

```

Qy 98 PAKAEEBAQSLF-----VTECI-----KTVNSDMHLVNYKYEDYSGCF 135
Db 436 PFKASREDSGLFSPRRSAFSPGLGCTPAFCFCQDIDGDIHENDHDSVYTYEDYA---- 492
Qy 136 RQLPNKVVKLDKLVPHVYEVDEVDKEDDASLSQKQGTITGGMVLYGNGNSAISVTWR 195
Db 493 -----KSISCEVLGSLVLRTHHTNTLSININSIKHGKNTVTPFGHNI,DOKNKSKNSKSLMIK 547

```

QY	196	SEKREF--PHLLOLGGGSKYFEFLKOLQOEKPSIFLGLVGSFNNKVRRAPELKNOD	254
Db	548	DSIQFAADVLVEKSGS-----AFKOLQK-----GVSECTNML-----	580
QY	255	KSSVLLAADSEVEEMEWITLNLKILQLNFEAMQEKRGDSHEDDEQSKLESGSGLDSTY	314
Db	581	---YHLA-----IKLTSSVLQMAFDELRROR-----AFSLKERAIIGLNF	618
QY	315	L--PELASAREAEIKGS--ESRYKLVYLLDDACKLOF-----SSAPEVK	357
Db	619	LVSEALSNALKDOLQYVKQIFLNTVAPRAADL--ABELVPEGIMEVQCPSPOTPAISPCCG	677
QY	358	SFEKFKGRILVCKNDLSFNLQCCVAENEBEGFTVWEPFVTLISFDIKYNRKISADHV	417
Db	678	SFDFE--DKVYVLYKADLSESV-----IQAFIELSGVDVTLTTKAAVSVST	722
QY	418	D-LNHFVSROMIATT-----SPALMNGSGPETQSLRGILHEAMQYR--KQGISVTCPH	470
Db	723	DNIKVVSSESVPSPSOAVTFSPSPFNQOIMYTKP-----VOEYKKEVYVOQALFCTS---	774
QY	471	PDIFLVARIIEKVLQOSI-----THC--ABPYMKSDDSK-----VAKVULKANKQ	513
Db	775	---GIVTSIPLPPLAGSALLPHYISSTAOAKKHLSDSDSNSGDSAQVHIATK--NREKA	830
QY	514	ACOR---LGOYRMPAPMAARTLFKDXASGNLDEKNARPSAIYRDSNKLSDNDMLKLLADR	570
Db	831	ACLNRNICLPSEHNP-----GN-----QNDKPKTNDI-----EMQ	860
QY	571	KPEKAKLPVLGNLDITIDNWSDFPVPVNVSSYIPQIOFECSTKIPTTFVEEVEPCIP	630
Db	861	SSSKLPDNPALISNSAAVHT-----IVNLTLSMTLSLEYTKWDETTDYLT	908
QY	631	KHTOPYTITYNHLVYVPKYULKYDSQSEFAKKNIAICIEFXDSDDEDSQPKCIYGRPG	690
Db	909	KSLEKTEPPFSH-----C-----DQAVIQCSBASANK	935
QY	691	PVFTRSAPAAVLHHQNPBFYDEIKETLEPTQHEKHLLLTFHVSCNNSKSGSTYKRDV	750
Db	936	DMFADRLSKSIHKH-----IDKSKSVIP-----NIDKNAYKES-	970
QY	751	VETQGVYMWLPLLKGRVVTSEQ--HIVSANLPBGYGYOELGWRHYGEIK--ATWDG	806
Db	971	-----LVSGSESQLTPEKSPKFPDSON-----QLTHCSLSAKQCVBECKASMYHG	1017
QY	807	G-----KPLK--ISTHLVSTVYTDQ--HLHNFQYQ	836
Db	1018	SSLFTLPCSPAVTGQKDLKESAKQPKLKNLNTSLEALSFGDENPFPSSHFTSSYAL	1077
QY	837	KTESGAQALGNELVYULKSLSHAMECHVNIAPLPTLNLQFLVLRATQOEVAANVTRII	896
Db	1078	TCVDDLHVEDKOKVR-----DKNV-----IPDTPSTPLVPSSASSEMDIKULTTKLK	1125
QY	897	HVVAAQ-----CHEEGLESHLSRYVKAAYKAPVASEYKVTHELTYSMTTILKP	946
Db	1126	GELAKEFAPATPPSPPHNVSQGS--LSBNKQNTIEKEBFMLKLMRSLSEVSSBEGEL--P	1183
QY	947	SADFLT--SNKLTK-----SWFFPVLIKSMQHL-----TENSKYK--LLNROR---	988
Db	1184	EVDVSEHSGKKVQOPEALATHILISLATEMASHSIDNKIOEPKQNDCLNWQSORSVSP	1243
QY	989	-FPASVYHAAVETVVV---MLMPHITQKFRDUPPEASKNANSLSAVFIKRCFTPMDRGVVF	1043
Db	1244	TFLWPSDNLKTLICNFADDLAEVLT-----ENAKIAK-----VANC-----MLF	1283
QY	1044	KQINNYISCFAPGDPKTLFEYKFEFLRVCNHEHYI--PLNDMPRG--KGRIORYODLO	1099
Db	1284	KOKKX--SCYADGEDYKVEEKLEDTIAVV--HPREVDFIISLPSSCMGSMGYPCSE	1339
QY	1100	LDVSLTEFCGNHFLVGLLREVGVALOPEFSEVRLIALSVLKNLILKHSPPDRKVASRHO	1155
Db	1340	---SVTDEYAGH--LIQILKQEGGS-----ELIM-----DOYANR--	1370

```

OY 1160 ARITLVLPFLGGLLENNQVINTV-DVSPFVNAGMTVDCESLAL-----1203
Db 1371 ----LAIRSVASGLQEAAKTKTCQCNRMFVFPVSQYKTKNELLMFSENKEHHQADKKRQ 1426
OY 1204 -----PAVNPVLTPQKGT-LDNSLHKD-----LGAISGIASPYTT 1239
Db 1427 SKRNEGVPCKNQTCERTLDPYRNREVSQLYSPSTSLVSHSITKDAKEELTASIVGLPKSLTD 1486
OY 1240 S-----TRINSVRNADRGSLISTDSGNSLPERNSEKSN 1275
Db 1487 SCLEFKSGYEEDNECHVTPPELPSKLOPSQNHRRYHSTGSLNGYGCDDNVQAVEQYAKK 1546
OY 1276 LDKHQQSSTLGSNVVRCDKLDQSEIKSLMCFYLILKMSMDALFTYWNKAS--TSELD 1333
Db 1547 VVDLTLELTLSQSTVFRVSETTKASDR-----VTTAEKLSPLTGQACR 1588
OY 1334 FFTISEVCLHFOFYMGKRYIARTGM-----HARLOQLGSLDNSLTFNHSYGH 1382
Db 1589 YCDLKE--LHNCGTNSSQHFFRQSLASKPKASPKPSRQK-----SRIPLHSVPQI 1640
OY 1383 DADVLHOSLEENATATEVCLTALDLSLFTLAFKQLADHGNHPLMKKQFVDVYLCPLQK 1442
Db 1641 HVNLDKAVLLEKIVAEALIEKARELESTSLA-----ADSG-----IQEAGASPAES 1687
OY 1443 HQSET---ALKNVFTALRSL--LYKPFSTPEGRADMCALCYEIL-KCCNSKUSIRTE 1496
Db 1688 LATETMTAVTNVHAHVSSKEISDFOSTESVSQQANLSTIGDSTGSWSVLSFEDEHD 1747
OY 1497 ASQLLYFLMRNPFYTGKSKSFVRTHLQVIISQLIDVAVGIGCTREQOQSLIINNCA- 1555
Db 1748 ESSSFHHISENGSSMS-----LGLBEDLYEDNISFTPSDSG 1788
OY 1556 -SDRLIKHTSFSSVYKDLTKRIRTVLMTAQMKENDEPEMLVDLYSLAKSVSTPEL- 1613
Db 1789 PDDKDEEH---EDVEGIGQDQKTLITINIDMECTVDPOLRIILQWLIA--SEAEVVELY 1844
OY 1614 -----RKTWLDSMALIHVNGLSERAMCYVHTVLAIVELYLRKXAVQ 1656
Db 1845 FHDSANKEFMLSKOLOEKGM-----KVGLDQAVIQYEVMEKASSEERCKSLFD 1895
OY 1657 W 1657
Db 1896 W 1896

RESULT 14
ACF7_MOUSE STANDARD; PRT: 5327 AA.
ID ACF7_MOUSE
AC AC Q9QXZ0; P97394; P97395; P97396.
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Actin cross-linking family protein 7 (Microtubule actin crosslinking factor) (MACF) .
DE ACF7 OR ACF7P.
OS Mus musculus (Mouse) .
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2) .
RC STRAIN=BALB/c;
RX MEDLINE=20069791; PubMed=10601340;
RA Leung C.L., Sun D., Zheng M., Knowles D.R., Ilem R.K.H.;
RT "Microtubule actin cross-linking factor (MACF): a hybrid of dyonin
RT and dystrophin that can interact with the actin and microtubule
RT cyoskeletons.";
RL J. Cell Biol. 147:1275-1286(1999) .
RN [2]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3) .
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=97124842; PubMed=8954775;
RA Bernier G., Mathieu M., De Repentigny Y., Vidal S.M., Kothary R.;

```


QY	350	SSABEYVSPBEKPEKRIILVKNCDNSFNLQCCVAAENEGPPTNNBPFFVTLSPDIXKNR	405	
Db	1594	KDLOQDQSHSTSPATAVK-----IIEGLEENQ-----T	1623	
QY	410	KISADPHVDLNFYSVROMIATTSPALNMGSGPETOALSRLIHEAMQYPKOJIFSUTCP	469	
Db	1624	KLS-----POEITLALREKHLQKQEGYE-----	1645	
QY	470	HPDITLVAR---IEVYLOGSITHCPEPIKMSDSKVAQKVLKAKQACQRLG-----	519	
Db	1646	---VLOERTVRAQOELEBAVTSALQ---QETEKSKATTELAENRKIDALLDWVGLMG	1698	
QY	520	---QYRMPF-AMA-----RTLFKDASGNLDGN-----ARFSAIYRQD	553	
Db	1699	HLRESHRLAFQRMSSLELAWEKOTLANTDGHVDVNOQYPETLDRQYELMKARHQELLSQ	1758	
QY	554	SNKL---SNDMLKTLADFPKKEPMATLPIYLGWLDITIDVSSDPFNVNSSTIPTYQ	609	
Db	1759	QNFIVATQSAOSFLDQHSNHLPEEROKLOEGLKELK-----EQYAAFS--LASE	1806	
QY	610	FETGSKDPTTFPEVEFEVPCIRPHQTPRYITNHLVYVPKYUKYDS-QKSPKARNIACI	668	
Db	1807	AEKQTOALRDELQKFL-----QDHKEFENWLOQSENEL--DSMHKGGSSPEALNSL	1855	
QY	669	EFKSDSEEDSOPKLKCIYRPGGPAVTRSAFAVLHMHQNP-----FYDEIK--	715	
Db	1858	KRQGSFSED-----VISHKGDLRFVITISQVLEETENNFBQGPSPATRLVMEKLD	1911	
QY	716	IELPTQLEK-----HHLLTF-----FHVSGDN-----SKSGTKR--RDVETQVG-	756	
Db	1912	TERYITLHLSKICIRLQSHLSMLIGYOQFOSSADSIQAWLTCBAVGLKSLDTVAASDGV	1971	
QY	757	-YSWPLRLKDGKRVVYSEOHIPYVSANLPBGVLGYOCELGMRHYGPRIKWNDSG----	KPL	810
Db	1972	LQOQATTKQJQOELEAHQVPP-----EKLOQKAH-----DLDDIEBPALDCRPI	2017	
QY	811	LKISTHVLVTVYTTQDQHLNHFQOYCOQTESGAQALGNELVYUKLSLHMEGHVMIAPLT	870	
Db	2018	QETDTSISSRFONLSCSIDERSALLQKAIASQOSQOESMESILQSIREVENOL-----	2070	
QY	871	ILNQLFRVLTRATQOEVAVNVNTRVIVHVAQOHEGLESLSHSYKYVYVKAEPYVASEYK	930	
Db	2071	-----ERQOVASSLSGVT-----QALANNMKLKODIARO-----KSLE	2105	
QY	931	TVHEELYSMTTILKPSADFLTNNKLYKSWFF-----FDVL	967	
Db	2106	ATHDWVTRFMEFADNSASAVL-QGKLAEISQFQOLQOQOEKESNLIKLLPQAEMLFQOL	2164	
QY	968	IKSMAOHLIENSKVULNLNQRFPASVYHNAVEIVVNMMLPHITQKRDMPKASKANSHLA	1027	
Db	2165	SNKLQOQFEMENSKRLASGNQPDQDIHFSQO-----IQEITLAMEDKEXMLDTLEH-LV	2217	
QY	1028	VFIKRCPTFMDRGFVFKQINNYISGFA-----PGDPKTLFPEYKKEFLRVYCNHBY	1078	
Db	2218	TTLGSCGFALDLSQHQDKIQNLKKQFTELQTVQERKEDASTQEQOJDEFPKLLRTFPQM	2277	
QY	1079	IPL---NLP--MPFQGR-----IQRYDDLOLDYSLTDFCRNHFVGLLIREV---GTA	1125	
Db	2278	LKETEGNVPKATFVSAKELEKQIEMHLDLISDESK-----GALLGEIINAKCTA	2327	
QY	1126	LOEFREVRLIAISVKNLKLKHSFPDRVAASSHORATIPLYLPGLLIENVOIRNV-RD	1184	
Db	2328	LE-----SLMDITAPDS-----QAKTGST-LPPVGSVGSVNGYHTTKD	2366	
QY	1185	VSP-----FPVNAQ-----MTVDESLAPAVNPLVTPQKSGTITDMSLH--KDLLGAI	1230	
Db	2367	LTEIQCMDFDVSNSKEKLEMLVLRHQESLQYFVSMEVQKXASAVLQWLSKEBEVLKAM	2426	
QY	1231	SGIASPYTST-----PRINSVRADSRGSLISTDSGNSLPERSKE	1271	
Db	2427	DATUSTYETETVAKQAESNKAFELAELEONSPTIQCVKA--LAGLLKTPYNSQEAENWK	2484	

QY	1272	KSNSLDKHOQSTTGNISVVRCDKDOSEIKSLMCFYLTKSMGSDALFTYVNAASTSEL	1331
Db	2485	MOEDLNSRMEKAT--EVTYAROKOLEES--ASHLACF-----	QAAASQL 2524
QY	1332	MDFFTISEVCL-----HOFYQKREYIARTGMHARLOOL--GSL--DNSTL	1373
Db	2525	RPMLEKEKLMMGVGLPSIDPNMLKQOVFLKFEARR--QHQRQLMBAAGILTGPGDM	2583
QY	1374	TFNHSYGHSDAVLHQSLEENANITEVCLTALDPLSLFLTAFLKQQLADHGHNLMKKVF	1433
Db	2584	SPSASOVHKDQGISOKKVE-----LT--DLNLSRSQIDQAIYKSTQYODLLODLS	2633
QY	1434	DVYLCEFLQKHQSEFALKNVFTALASLLYKPFSTYEGRADWCALCYEILKCNCKLSKI	1493
Db	2634	EKVATIGORLUSGSAISTQPEAVVQOL-----EETSEI	2666
QY	1494	RTEASOLLYFLMRNPNFYTGKKSFWRTHLQVLIISQVLADVVGIGSTRFOQSLSIINN	1553
Db	2667	RSDDIGQL-----DNEIKEA-----QTL-----C	2684
QY	1554	ANSRLIKHNSFSSSDVKDLTKRITFTYLMATQMKENHENDPEMLVDLOYSLAKSYASTPEL	1613
Db	2685	QELSLLIGEQYLKDELK--KRETFVVLPLQGLE-----DLADMRSLQAPALASTQOF	2735
QY	1614	R-----KTMIDSMARLHYKNGDS---EAMCVYVHTALVAELTRKENVQMEPRLPH	1664
Db	2736	QOMPELRTKLTDEKQSOQAQKCPISAKLERLOCL-----QENEEFKALNNQH	2783
QY	1665	SHSACLRRSRGVFRQCTAFRVITPNIIDEBASMMEDVGMQDVAFNEVDLMELL---EQC	1721
Db	2784	SGSEIVYVABE-----EALLSVPRGEEKTKLQNLVLRSGW--BDLSKTKTANRQSRL	2835
QY	1722	ADGLMKAREVELADIYKLIIPYE--KRDFERL---AHLVTLHRAVSKTEVMSHG	1775
Db	2836	KDCMQAKQKQY--GHVEDLVPWIDECKSKMPELQVTLDPVQLESLLRSKAMLNDA--EK	2891
QY	1776	RRLIGTFPRVAFFGQAQYQPTDETFVEGFEBEDG---KEIYYE--PKLTPLSEIS	1829
Db	2892	RRSL-----LEINLSADILINSEIDEDERDEKAGINOMDAITBELQAKTSLEMT	2946
QY	1830	QRLKLKLYSDKGSENVMMIODSGVNPBKIDSKAYIQTVHPIDPFP-----EKELQ	1881
Db	2947	QRL-----KEPQESFNRIEKRYBG-----AKHQLEIFDALSGQACSNKYLE	2987
QY	1882	ERKTEFE-----RSHNIRRFMEFMTOTGKRQGVVERQCKRITLLIHNCPYVKR	1934
Db	2988	KLKAQOEVLQALBPQVYLRNF-----TQGLVEDA-----PDGSDA	3023
QY	1935	IPWVYQHTDNLNPLEVLAIDEMSKVAELRLQCSSAEVDMTKLQKLQSGSVSVQVNAAPLIA	1994
Db	3024	SPLVHQ-----AEVAQOEF-----LEVKORVSS---CLTMEKTLGIGQFHCVRBEMF	3069
QY	1995	YARAFLODPT-----NTRKVPDNKKVTKLEVFQRCVBAAGQALANBERLIKEDQLE---	2045
Db	3070	SQLADLDELDGMAIGRDTDSLOSQIEDV--RLFLNKL--QALRPD---IEDSEAECKML	3124
QY	2046	QENMKANYREMAKELSEIMHEQICPLBEKTSVLPNSLHI	2084
Db	3125	EEEGTLDLILGUKREL--EALNKKQCGKLTNRGCVVRQEQLEL	3162
RESULT 15			
YD86 SCHPO			
ID	YD86 SCHPO	STANDARD;	PRT; 1957 AA.
AC	Q10411;		
DT	01-OCT-1996	(Rel. 34, Created)	
DT	01-OCT-1996	(Rel. 34, Last sequence update)	
DT	15-JUN-2002	(Rel. 41, Last annotation update)	
DE	Hypothetical protein CJF3_06c in chromosome I.		
GN	SPAC1F3_06C		
OS	Schizosaccharomyces pombe (Fission yeast).		
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;		
OC	Schizosaccharomycetales; Schizosaccharomycetaceae;		

CC Schizosaccharomyces.
 OX NCBI TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RK MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Galliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 RA Sgouros J., Pat N., Hayles J., Baker S., Baeham D., Bowman S.,
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 RA James K., Jones L., Jones M., Leather S., McDonald S., Mclean J.,
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., O'Neill C.,
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 RA Taylor J., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 RA Woodward J., Volckaert G., Aert R., Robben J., Grynopreuz B.,
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fricz C., Holzer E., Moestl D., Hilbert H.,
 RA Borzym K., Zimmermann W., Medler H., Wandt R., Reinhardt R., Pohl T.M.,
 RA Egger P., Zimmermann W., Medler H., Wandt R., Reinhardt R., Pohl T.M.,
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 RA Galbert P., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
 RA Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 RA Dominguez A., Revelta J.L., Moreno S., Armerstrong J., Forzburg S.L.,
 RA Cerretti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 RA Shpakovsky G.V., Useery D., Bartell B.G., Nurse P.;
 RT "The genome sequence of Schizosaccharomyces pombe";
 RL Nature 415:871-880 (2002).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: Z70690; CAAB4624.1; -
 DR Hypothetical protein.
 KW SEQUENCE 1957 AA; 222785 MW; 3F480CA06171D9DA CRC64;
 SQ
 Query Match 1.7%; Score 181.5; DB 1; Length 1957;
 Beest Local Similarity 17.4%; Pred. No. 0.035;
 Matches 369; Conservative 326; Mismatches 803; Indels 623; Gaps 83;
 7 LPASATRTKFTALSKGTAELROSVEVVRGSVLLAKPKL-IEPLDYENV----- 57
 176 LTSEEDVSYFOKKLTMTESNPFASKOSEAIDLROLITVTKDKKEDYETIKEDVSSIK 235
 58 IVOKTQIINDCLR-----EMLLFPYDFOFALIRGQ--RYICSTVPAKAE---BA 105
 236 ASLAEQANSKSLRGEOERLEKLVSSNKTIVSTLRQTENSLRPECKTLQDEKLEKAINB 295
 106 OSLPTECICKYTNMDMLVNYK-----YEDYSGFRQLPKNKYKDKLPVAYEVD 156
 236 DSKLLEEL-----KINVANYSDAIVHKDKLIEDLSTRISFEDNMLKSEDDTISIKNEKE 349
 157 BEVDKEDDASLGOKGITTGKGMVLKGMNSAISTVTRSPK--RRFPLILQGDSDSYF 214
 350 KLLRN-----TIGSLKDSRTS-----NSOLEBEMVLKESNRTIH-SQITDSESL 394
 215 EFLKDLQKSPKGSIFLGLYGVSPNNKVRPAFAFLKMDKSSYLLAADSEVEMEWITI 274
 395 SSFPEBNKSLKGSID-EYONNLSSKDKVMYKOVSSQLE-EARSSLAHATGKLAINEERDF 452
 275 LKKIILQNEEAMQER---NGDSHEDDROSKL-----EGSGS 309
 453 QNKKIK-DEKTIQDRLACLSSSNELKERSALIDKQDELNNLRQIKQKKVSESTOS 511

QY 310 GLDSYLPFLANLSAREAEI-----KLKSESRVKLFYLPDPDAOKLDFSSABPEV----- 356
 DB 512 SLQSLQDIILNKKKHEVYESQLNELKQELQTEISNSBLSLSQSLTAAKKAIVATNNE 571
 QY 357 -----KSEEFKGRILVKKNDLSFNLOCCVAENBEGTTVVPFVLSLFD 404
 DB 572 LSEKNSLQTLGNAPQOEKLASVWQ-----LKEBQ-----NFSILD 608
 QY 405 IKYNEKISADPFVNDHSPVROMITTSPALMNGSGPETOSALRGILHEAMQYKQIGF 464
 DB 609 TSFKLNSHQELENHNOITIKQLDTS-----SKLOQLQERANPEQKESTL 656
 QY 465 S-----VTCEHPDIFVARIKYV-LOGSITHCAEPMYKSSDSSKYAOKYLNKAK 512
 DB 657 SDENNLDRTKLLKEESKSLIKKQEDVDSLEKNIQTLKEDLRKSEBALRFSKLAQULR 716
 QY 513 QACQRL-QQYMPFPMARLTFKASGULDKRARSAIYRQDSNLSDDMLKLADF-- 569
 DB 717 EVIDMLKGRHE-----TLEQRNDLHSLSDAKNTALISELTQ-SEEDVRLTRANVET 770
 QY 570 --RKPEKAKLPIVIGNDITIDVNSDF-PRVYVSSYIPTKQPEKSKTPTFPAVEBFV 626
 DB 771 LTQDSKANKQSTLSVNSYQISINLYHELRDDHV-----MOSQNTLLESSEKLT 822
 QY 627 PCIPHTQPYTYITNHLVYPRYLKYDQSKSPAKARNIACIBFQDSDEDSQPLKITYG 686
 DB 823 DCENLTQGNMTLIDVQGLMKHKNVQSKVSELKEVNGKLSIDLK----- 868
 QY 687 RFGGVPPRSPAAVLHHQNPEDYDIKIELPTQJHEKHHLLTPFHVSCNDSKSGSTK 746
 DB 869 -----RRSSINVAISDND-----QILTQJLAE-----LSK 892
 QY 747 KRDVETQVGSWMLDLNDRGVTSQHIPIVSANLPSGYLGVELGMRHGYEPIKMWG 806
 DB 893 NYDSIQE-----SAQLNSG-----IKSLBA 913
 QY 807 GKPLIKISTHVLVTVTODQHL---NFOYQCKTESGAQALGNEL-----VYLYKSL 856
 DB 914 EKQLL-----HTNEBELHRLDLTGTLKIEBSKSDGLKTLARQEIENLKEB 963
 QY 857 HAMEGHVIAFLPTLNLQFLVLTATQGEVAVNVTIVYIIVVAOCHEBGLSHLSRYVK 916
 DB 964 NMSQQA-----ITSYKSLDETLSKSKSLKEDIE-----HLKNVSEVEVERNA----- 1008
 QY 917 YAYKAPYVASEYKTVHBEITKSMWTLIKPSA--DELTNKLKXSWPPFDLI----- 968
 DB 1009 -LLASNERLMDLKNNGENIASLQTEIKGRBENDLOS-KLSVSVSEYENLLISSQTN 1066
 QY 969 -----KSNAOHLIENSRYKL--RNQRPASVYHNAVETVVMMLMPHIYQKPRDNEASKN 1021
 DB 1067 KSLBEKTNQKTIKNGVQKLDEKQD-----HVELEBELTSKYQKGEENAO 1113
 QY 1022 ANHSLAVFIK-----CFTPMDRGVFVKQIINNTISCFAPADPRTLBYKPEPLRVN 1074
 DB 1114 IKDELLARKKSKQHQDLCANFVD-----LKEKSDALEOULTN 1151
 QY 1075 HEHYFLMLPMPFGGR--IQRYQD-----QLOVSLTD-----EPCRNPIVGL 1117
 DB 1152 EKNEELIVLEQSNNSNEALVEERSDLNRLSDMKSLSDSDNSVISIRSDLVRAVDELT 1211
 QY 1118 LLREVGTALOEFREVRLLAIVLKLIL-IKHSFDRVARSQHARIVTLVPLFILLTEN 1176
 DB 1212 LKQKDSISTGYSCQDRDLDBLKKCEBSFNKYNASLABELCTKSHIDVVSIELDN 1271
 QY 1177 VQRIINRVDSPPVNAQ-----MTVQESIALPVPVLTPOKGSTLDNSL----- 1222
 DB 1272 -----FVFAGNFSELSRLTVLSLNVYDAFQV--NFKKMEIDNLTJTTDAEF 1318
 QY 1223 -----HNDLGAISGLASPYTTS-----TPNHSVRNADR----- 1253
 DB 1319 TKVVADLEKLOHEHDBDWLQRGDLEKALKDSBKQFLRKAEAMTENIHSLBGKETTKET 1378
 QY 1254 GSLISTDSGNSLPERNSEKSNLSLDKQOOSTLGNSVVRCDK-----LDQSEIKSL 1303

```

Db 1379 AEISSRLBNOL--ATNKLGNOLDHNLQEI RLKEDVLKESLIIISLESISNORQES8 1436
Qy 1304 LMCFLYILKSMDDALFTYWNKASTSELMDFITISEVCLHQFOYMGRIYARTGMHARL 1363
Db 1437 LLDKXBLEHMLDD-----TSRKNSSLME-----KTESINSLDDKSFELASAV 1480
Qy 1364 OOLGSLDNLSTFNHSGHSDADVLHOSLEANIATEVCLALDPLSLFTLAFRN----- 1417
Db 1481 ERLGALOKL---HSESLSIMENIKSLOLEAKKEKIQVDESTIQELDHETIASKNNEGKL 1536
Qy 1418 ----QLADHGHNP LMKVFDVYLCFLQKHQSTALKNVFTALRSLIYKFPSTFYEGRAD 1473
Db 1537 NDRDSIIRDLSEN--IEOLNNLLA-----EERSAVKRLSTEKESILQF----- 1578
Qy 1474 MCAALCEYLKCNLSIRTEASOLLYFLMNNNDYTGKSFVTHLOVIIISVQLTA 1533
Db 1579 -----NSRLADLEYHKSOV-----ESELGRSKLKLASTTEEL-- 1610
Qy 1534 DVVGIGTRFOQSLISINNCAUSDRLIKHSPSSDYVDLTKRIRTVLMATAQKHEHNDP 1593
Db 1611 -----QAEENERL-----SLTRMLDLQNOVKDLSNIKDLSLEDLRTL 1648
Qy 1594 EMLVDIYSLAKSYASTPELRKTWLDSMARIHVKNGLSEAMCYVHTALVAEYLTRKE 1653
Db 1649 RSLIEDSVASLOKCKIKSNTVESLODVLTSVQARNALBD-----EVSRSYDKIRRDD 1702
Qy 1654 AVQWBPPLPHSHSACLRSGRGVFR---QGCTAFRVTFNIDEASMMEDVGMODVHF- 1709
Db 1703 RCEHLSGKLLKLSQ--LEBQHETFFRAEQORMTQLOGLKETVKKQKLLKLANRQOL 1761
Qy 1710 -NEDVIMELLEOCADGLMKAERYE-LIADIYKLIPIYERKRDPERLAHLVDTLHAYSK 1767
Db 1762 PRSSIIY-----YESYIRDIKEKIIIVLOERLNGIE---LSQOLPKGY-- 1800
Qy 1768 VTEVMSGRLLGTFRV-----AFPGQAQYQFTDSETDVGEFPEDEBGEYIYK 1818
Db 1801 -----FGYFFKTRNREMEVLDSFKQYAKLOFLAGAEFIYKFKEDLEKCAAEK 1849
Qy 1819 BPKLT-----PLSEISORLLKLY-----SDKFGSENVKMIODSGKYNPKD 1858
Db 1850 EKQATFDNTSEKVENIGKSI EALYFALNREISFRKSLALSKSAYHNL-LVRDSPKFPN-- 1906
Qy 1859 LDSKVAYIQVTHVIPFEDEKE 1879
Db 1907 -DS-----QITYSIPVTNTRQ 1921

```

Search completed: July 14, 2003, 18:16:35
 Job time : 46.9153 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:07:56 ; Search time 97.7492 seconds

(without alignments)
4441.379 Million cell updates/sec

Title: US-09-815-379-8

Perfect score: 10936
Sequence: 1 MSQPLPLPASAEKTRKFTPL.....ISGRPTSTWVGMTSSSSSV 2107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8002	73.2	1539	4 Q9B229	Q9B229 homo sapien
2	4605.5	42.1	909	4 Q9B227	Q9B227 homo sapien
3	4491	41.1	878	4 Q9B228	Q9B228 homo sapien
4	3693	33.8	724	4 Q9B226	Q9B226 homo sapien
5	3305.5	30.2	1984	5 Q9VM04	Q9VM04 drosophila
6	2888	26.4	738	11 Q63603	Q63603 rattus norv
7	2739.5	25.1	881	4 Q96MN3	Q96MN3 homo sapien
8	2192	20.0	448	4 Q9B225	Q9B225 homo sapien
9	2180	19.9	611	4 Q8WX10	Q8WX10 homo sapien
10	2094	19.1	417	11 Q921Y6	Q921Y6 mus musculu
11	2085.5	18.4	425	6 Q9GM45	Q9GM45 macaca fasc
12	2007	18.2	696	4 Q9BWX9	Q9BWX9 homo sapien
13	1985.5	18.2	534	4 Q8WX11	Q8WX11 homo sapien
14	1977.5	18.1	2374	5 Q45377	Q45377 caenorhabdi
15	1887	17.3	1538	4 Q9H1Q3	Q9H1Q3 homo sapien
16	1887	17.3	1540	4 Q9H1Q2	Q9H1Q2 homo sapien

17	1785	16.3	542	4 Q96BY6	Q96BY6 homo sapien
18	1759.5	16.1	2284	5 Q8TIC8	Q8TIC8 dictyosteli
19	1733	15.8	500	4 Q9NIX8	Q9NIX8 homo sapien
20	1680.5	15.4	1782	5 Q9VPI9	Q9VPI9 drosophila
21	1664.5	15.2	2018	5 Q20487	Q20487 caenorhabdi
22	1656	15.1	1373	4 Q8TEP1	Q8TEP1 homo sapien
23	1625.5	14.9	1628	4 Q9P2F2	Q9P2F2 homo sapien
24	1593.5	14.6	1302	4 Q9C092	Q9C092 homo sapien
25	1523	13.9	1180	4 Q9H7P2	Q9H7P2 homo sapien
26	1492.5	13.6	2621	5 Q8SSW5	Q8SSW5 dictyosteli
27	1395	12.8	415	4 Q9VM06	Q9VM06 homo sapien
28	1230.5	11.3	860	4 Q96HP0	Q96HP0 homo sapien
29	1226	11.2	595	4 Q75178	Q75178 homo sapien
30	1188	10.9	844	11 Q9CUB3	Q9CUB3 mus musculu
31	1141	10.4	624	4 Q8TB82	Q8TB82 homo sapien
32	1128	10.3	626	4 Q96NG5	Q96NG5 homo sapien
33	1084	9.9	444	6 Q9BE49	Q9BE49 macaca fasc
34	1043	9.5	1830	10 Q8SAB7	Q8SAB7 arabidopsis
35	1034	9.5	567	11 Q8RLA4	Q8RLA4 mus musculu
36	975	8.9	569	11 Q8VDR9	Q8VDR9 mus musculu
37	945	8.6	513	4 Q8WUY2	Q8WUY2 homo sapien
38	941	8.6	741	4 Q96N67	Q96N67 homo sapien
39	935	8.5	513	11 Q9DBQ2	Q9DBQ2 mus musculu
40	610.5	5.6	326	4 Q9H308	Q9H308 homo sapien
41	549.5	5.0	632	4 Q96N10	Q96N10 homo sapien
42	540	4.9	1105	10 Q23479	Q23479 arabidopsis
43	516	4.7	234	5 Q9VM06	Q9VM06 drosophila
44	429	3.9	1781	5 Q17758	Q17758 caenorhabdi
45	390	3.6	1907	4 Q15017	Q15017 homo sapien

ALIGNMENTS

RESULT 1	Q9B229	PRELIMINARY	PRT	1539 AA.
ID	Q9B229	Q9UPU4		
AC	Q9B229	Q9UPU4		
DT	01-JUN-2001	(TREMBLrel. 17, Created)		
DT	01-JUN-2001	(TREMBLrel. 17, Last sequence update)		
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
DE	BAL55N3.2.1 (KIAA1058 protein) (Fragment).			
GN	BAL55N3.2 OR KIAA1058.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RL	Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 6-1539 FROM N.A.			
RC	TISSUE=BRIN;			
RX	MEDLINE=99397452; PubMed=10470851;			
RA	Kikuno R., Nagase T., Ichikawa K., Hirosewa M., Miyajima N., Tanaka A., Kocant H., Nomura N., Ohara O.;			
RA	"Prediction of the coding sequences of unidentified human genes. XIV. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro."			
RT	DNA Rec. 6:197-205(1999).			
DR	EMBL; AL61420; CAC27814.1; -			
DR	EMBL; AB028981; BAA03010.1; -			
FT	NON TER			
FT	SEQUENCE 1539 AA; 176392 MW; 354F93C1120EBE71 CRC64;			
Qy	Query Match	73.2%; Score 8002; DB 4; Length 1539;		
Qy	Best Local Similarity	99.9%; Pred. No. 0;		
Qy	Matches 1537; Conservative	1; Mismatches 0; Indels 0; Gaps 0;		
Db	531 TLFKDASGNDLKNARFSAIYQDSNKLSDNMLKLLADFRKPEKAKLPVILGNLDTTD 590			
Db	1 TLFKDASGNDLKNARFSAIYQDSNKLSDNMLKLLADFRKPEKAKLPVILGNLDTTD 60			

```

QY 591 NVSSDPNNVNSSYIPTKQFETCSKPTITFEVEBEFVPCIPKHTQPTIYTNNHLYVYPKYL 650
DB 61 NVSSDPNNVNSSYIPTKQFETCSKPTITFEVEBEFVPCIPKHTQPTIYTNNHLYVYPKYL 120
QY 651 KYDSQKSPAKARNIAICIEFKDSDEBDSQPLKCIYGRPGGVPVTSAPAAVLAHHQNPBF 710
DB 121 KYDSQKSPAKARNIAICIEFKDSDEBDSQPLKCIYGRPGGVPVTSAPAAVLAHHQNPBF 180
QY 711 YDEIKELPTQHEKHHLLTFPHVSCDSSKSTKRDVETQVGYSMPLPKDGRVYT 770
DB 181 YDEIKELPTQHEKHHLLTFPHVSCDSSKSTKRDVETQVGYSMPLPKDGRVYT 240
QY 771 SEOHIPVSNLPSGYLOELGMGRHYGPEIKWVDGKPLKISTHVSVTYTOQHLYN 830
DB 241 SEOHIPVSNLPSGYLOELGMGRHYGPEIKWVDGKPLKISTHVSVTYTOQHLYN 300
QY 831 PFQYCOKTESGAQALGNELVKYLSLHAGHVMIAFLPTILNQLFRVLTTRATQEEVAVN 890
DB 301 PFQYCOKTESGAQALGNELVKYLSLHAGHVMIAFLPTILNQLFRVLTTRATQEEVAVN 360
QY 891 VTRVITIHVVAQCHEEGLSHLSYVYKAYKAEVYVASEYKTVHEELTKSMITILKESADF 950
DB 361 VTRVITIHVVAQCHEEGLSHLSYVYKAYKAEVYVASEYKTVHEELTKSMITILKESADF 420
QY 951 LTSNKLKYSWFFDVLKISMAOHLIENSKVKLLRQRPASVYHAYETVYNNMLMPHIQ 1010
DB 421 LTSNKLKYSWFFDVLKISMAOHLIENSKVKLLRQRPASVYHAYETVYNNMLMPHIQ 480
QY 1011 KRDNPEASKANANSLAVPIKRCFTPMDRGFVFKQINNYISCPAPDPTLPEYKFEPLR 1070
DB 481 KRDNPEASKANANSLAVPIKRCFTPMDRGFVFKQINNYISCPAPDPTLPEYKFEPLR 540
QY 1071 VVCNHEHYIPLNPMFPGKRIQYODLQDLSLTPEPCRNHLYVLLREVTALQERR 1130
DB 541 VVCNHEHYIPLNPMFPGKRIQYODLQDLSLTPEPCRNHLYVLLREVTALQERR 600
QY 1131 EYRLIAISVYKULLIHSFDRYASRSHOARITATLPLFGILLIENVQINVDVSPFV 1190
DB 601 EYRLIAISVYKULLIHSFDRYASRSHOARITATLPLFGILLIENVQINVDVSPFV 660
QY 1191 NAGMTYKDESLALPAVNPVTPQKGSITLNSLHKDILGAIISGIASPYTTSTPNINSVNA 1250
DB 661 NAGMTYKDESLALPAVNPVTPQKGSITLNSLHKDILGAIISGIASPYTTSTPNINSVNA 720
QY 1251 DSRGSLISTDSGNSLPERNSEKSNSLDKHQOSSTIGNSVVRCDKLDQSEIKSLMCEFLYI 1310
DB 721 DSRGSLISTDSGNSLPERNSEKSNSLDKHQOSSTIGNSVVRCDKLDQSEIKSLMCEFLYI 780
QY 1311 LKMSMDALFTYNNKASTSELMDFITISEVCLHQFQYMGKRYIARTGMHARLQOLGSLD 1370
DB 781 LKMSMDALFTYNNKASTSELMDFITISEVCLHQFQYMGKRYIARTGMHARLQOLGSLD 840
QY 1371 NSLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDITSLFTLAFNQOLLADHGNPLMK 1430
DB 841 NSLTFNHSYGHSDADVLHOSLLEANIATEVCLTALDITSLFTLAFNQOLLADHGNPLMK 900
QY 1431 KVPDYVLCLOHQOSTALKNVFTALRSIYKFPSTFYEGRADMCALCEYLKCNSTKL 1490
DB 901 KVPDYVLCLOHQOSTALKNVFTALRSIYKFPSTFYEGRADMCALCEYLKCNSTKL 960
QY 1491 SSIRTEASOLLYFLMNNFDTYTKSGFVETHLOVIVISVOLADVVGIGTFEQOGLSTII 1550
DB 961 SSIRTEASOLLYFLMNNFDTYTKSGFVETHLOVIVISVOLADVVGIGTFEQOGLSTII 1020
QY 1551 NNCANSRLIKITSSSDYKDLTKRIYTLVMTAQKHEHNDPEMLVDLOYSLAKSYAST 1610
DB 1021 NNCANSRLIKITSSSDYKDLTKRIYTLVMTAQKHEHNDPEMLVDLOYSLAKSYAST 1080
QY 1611 PELRKTWLDMSMARIHVKNGLSBAAMCYHVTALVAEYITRKAQVWEPPLPHSHSACL 1670
DB 1081 PELRKTWLDMSMARIHVKNGLSBAAMCYHVTALVAEYITRKAQVWEPPLPHSHSACL 1140

```

```

QY 1671 RRSRGVFRQCTAERVITPNIDEASMEDEVMQDVHFNEDVLMELLEQCADGLMKAER 1730
DB 1141 RRSRGVFRQCTAERVITPNIDEASMEDEVMQDVHFNEDVLMELLEQCADGLMKAER 1200
QY 1731 YELIADIKLITPIYEKRDPERLALYDTLHRAVSKYTEVMSGRLLGTYFRVAFQO 1790
DB 1201 YELIADIKLITPIYEKRDPERLALYDTLHRAVSKYTEVMSGRLLGTYFRVAFQO 1260
QY 1791 AAQYFTSEPTVEGFEDDEKEXIYKPKLTPISLISORLLKYSDPKFGEENVKMIOD 1850
DB 1261 AAQYFTSEPTVEGFEDDEKEXIYKPKLTPISLISORLLKYSDPKFGEENVKMIOD 1320
QY 1851 SGKVPKLDISKVAYIOTVHVPFDEKELQERKTEPERSHNIIRPMEEMPTQTKGQ 1910
DB 1321 SGKVPKLDISKVAYIOTVHVPFDEKELQERKTEPERSHNIIRPMEEMPTQTKGQ 1380
QY 1911 GVEEOCKRTITTAHCPEPYKRIPWYQHTDLNPIEVAIDENSKYAEIROLCSNAE 1970
DB 1381 GVEEOCKRTITTAHCPEPYKRIPWYQHTDLNPIEVAIDENSKYAEIROLCSNAE 1440
QY 1971 VDMIKLOLQKQSVSVQVNAAGPLAYARAFLODPTNTKXYPDNKVKLLKEVFRQFVACGQ 2030
DB 1441 VDMIKLOLQKQSVSVQVNAAGPLAYARAFLODPTNTKXYPDNKVKLLKEVFRQFVACGQ 1500
QY 2031 LAVNERLIKEDQLEJOEEMKANYREMAKELSEIMEQI 2068
DB 1501 LAVNERLIKEDQLEJOEEMKANYREMAKELSEIMEQI 1538

```

```

RESULT 2
Q98227
ID Q98227 PRELIMINARY; PRT; 909 AA.
AC Q98227;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BAI55N3.2.2 (KIAA1058) (Fragment).
GN BAI55N3.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Smith M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL161420; CAC27816.1; -.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 909 AA; 103445 MW; 77166F77937339F3 CRC64;

```

```

Query Match 42.1%; Score 4605.5; DB 4; Length 909;
Best Local Similarity 97.5%; Pred. No. 2,5e-285;
Matches 886; Conservative 0; Mismatches 0; Indels 23; Gaps 1;

```

```

QY 531 TLFKASGNDKNAFSAIYRDSNKLNDMDKLADFRKPEKAKLPVILGNDITTD 590
DB 1 TLFKASGNDKNAFSAIYRDSNKLNDMDKLADFRKPEKAKLPVILGNDITTD 60
QY 591 NVSSDPNNVNSSYIPTKQFETCSKPTITFEVEBEFVPCIPKHTQPTIYTNNHLYVYPKYL 650
DB 61 NVSSDPNNVNSSYIPTKQFETCSKPTITFEVEBEFVPCIPKHTQPTIYTNNHLYVYPKYL 120
QY 651 KYDSQKSPAKARNIAICIEFKDSDEBDSQPLKCIYGRPGGVPVTSAPAAVLAHHQNPBF 710
DB 121 KYDSQKSPAKARNIAICIEFKDSDEBDSQPLKCIYGRPGGVPVTSAPAAVLAHHQNPBF 180
QY 711 YDEIKELPTQHEKHHLLTFPHVSCDSSKSTKRDVETQVGYSMPLPKDGRVYT 770
DB 181 YDEIKELPTQHEKHHLLTFPHVSCDSSKSTKRDVETQVGYSMPLPKDGRVYT 240
QY 771 SEOHIPVSNLPSGYLOELGMGRHYGPEIKWVDGKPLKISTHVSVTYTOQHLYN 830

```

```

Db 241 SEOHIPVSNANLPSGLYGOELGMRHYGPEIKWDGKPLKISTHLYSTVYTDODHLN 300
Qy 831 FFQYCKTESGAQALGNELVYKLSLAMEGHVMIAPLPTLNOLFRVLTATOEBAVN 890
Db 301 FFQYCKTESGAQALGNELVYKLSLAMEGHVMIAPLPTLNOLFRVLTATOEBAVN 360
Qy 891 VTRVLIHVAOCHEGESHLRSYKAYKAPYVASEYKTVHBEITKSMITLKPSADF 950
Db 361 VTRVLIHVAOCHEGESHLRSYKAYKAPYVASEYKTVHBEITKSMITLKPSADF 420
Qy 951 LTSNKLKYSWFFPDVLKISMAOHLIENSKVLLRNORFPASYHNAVEVVMMLPHITQ 1010
Db 421 LTSNKLKYSWFFPDVLKISMAOHLIENSKVLLRNORFPASYHNAVEVVMMLPHITQ 480
Qy 1011 KFRDPEASKANANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDPKTLPEYKEFLR 1070
Db 481 KFRDPEASKANANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDPKTLPEYKEFLR 540
Qy 1071 VVCNHEHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHPLVGLLRREVGTALQEFR 1130
Db 541 VVCNHEHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHPLVGLLRREVGTALQEFR 600
Qy 1131 EYRLAISVLKXLLIKHSFDDRYASRHOARIATLYPLFGLLIENVORINRVDSPPV 1190
Db 601 EYRLAISVLKXLLIKHSFDDRYASRHOARIATLYPLFGLLIENVORINRVDSPPV 660
Qy 1191 NAGMTVKDESLALPAVNPPLVTPQKSTLDSNLHKDLGAI SGIASPYTTSTPNINSVNA 1250
Db 661 NAGMTVKDESLALPAVNPPLVTPQKSTLDSNLHKDLGAI SGIASPYTTSTPNINSVNA 720
Qy 1251 DSRGSLISTDSGNSLPERNSEKSNLSDKHQOQSTLGNSVVRCDKLDGSEIKSLMCFLYI 1310
Db 721 DSRGSLISTDSGNSLPERNSEKSNLSDKHQOQSTLGNSVVRCDKLDGSEIKSLMCFLYI 780
Qy 1311 LKSMDDALFTYWNKASTSEIMDFPTISEVCLHQPQYMKRYIARNOEGLAGIIVDRKSQ 1354
Db 781 LKSMDDALFTYWNKASTSEIMDFPTISEVCLHQPQYMKRYIARNOEGLAGIIVDRKSQ 840
Qy 1355 -----RTGMHARLQOGLSDNSLTFPHSYGSHDADVLHOSLEANIATEVCLTALDT 1407
Db 841 TLPVSRNRIGMMHARLQOGLSDNSLTFPHSYGSHDADVLHOSLEANIATEVCLTALDT 900
Qy 1408 LSLFTLAFK 1416
Db 901 LSLFTLAFK 909

```

RESULT 3

```

Q9B228 PRELIMINARY; PRT; 878 AA.
ID Q9B228;
AC Q9B228;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BA15SN3.2.3 (KIAA1058) (Fragment).
GN BA15SN3.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Smith M.;
RL Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL, AL161420; CAC27815.1; -.
FT NON_TER 1
FT TER 878
SQ SEQUENCE 878 AA; 99777 MW; C9475D4FE68B7231 CRC64;

```

Query Match 41.1%; Score 4491; DB 4; Length 878;
 Beef Local Similarity 97.4%; Pred. No. 4, 9e-278;
 Matches 863; Conservative 6; Mismatches 9; Indels 8; Gaps 1;

```

Qy 531 TLPKASGNLDKNAKFSAIYRQDSNKLNDMDKLADFRKREKAKLPIVLGNLDTID 590
Db 1 TLPKASGNLDKNAKFSAIYRQDSNKLNDMDKLADFRKREKAKLPIVLGNLDTID 60
Qy 591 NVSSDFPNYVNSSYIPTQEFCTCKPTTEVEEFPVPCIPGHOTPYTYTHLYYPRYL 650
Db 61 NVSSDFPNYVNSSYIPTQEFCTCKPTTEVEEFPVPCIPGHOTPYTYTHLYYPRYL 120
Qy 651 KYDSQSKAKARNIACIEFKDSDEBDSQPKCIYGRGSPVFRSAPAAVLHHNDEF 710
Db 121 KYDSQSKAKARNIACIEFKDSDEBDSQPKCIYGRGSPVFRSAPAAVLHHNDEF 180
Qy 711 YDEIKIELPTOLHEKHILLTFPHVSCNSKSGTKKRDVETQVGYMLPLKXGRVYT 770
Db 181 YDEIKIELPTOLHEKHILLTFPHVSCNSKSGTKKRDVETQVGYMLPLKXGRVYT 240
Qy 771 SEOHIPVSNANLPSGLYGOELGMRHYGPEIKWDGKPLKISTHLYSTVYTDODHLN 830
Db 241 SEOHIPVSNANLPSGLYGOELGMRHYGPEIKWDGKPLKISTHLYSTVYTDODHLN 300
Qy 831 FFQYCKTESGAQALGNELVYKLSLAMEGHVMIAPLPTLNOLFRVLTATOEBAVN 890
Db 301 FFQYCKTESGAQALGNELVYKLSLAMEGHVMIAPLPTLNOLFRVLTATOEBAVN 360
Qy 891 VTRVLIHVAOCHEGESHLRSYKAYKAPYVASEYKTVHBEITKSMITLKPSADF 950
Db 361 VTRVLIHVAOCHEGESHLRSYKAYKAPYVASEYKTVHBEITKSMITLKPSADF 420
Qy 951 LTSNKLKYSWFFPDVLKISMAOHLIENSKVLLRNORFPASYHNAVEVVMMLPHITQ 1010
Db 421 LTSNKLKYSWFFPDVLKISMAOHLIENSKVLLRNORFPASYHNAVEVVMMLPHITQ 480
Qy 1011 KFRDPEASKANANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDPKTLPEYKEFLR 1070
Db 481 KFRDPEASKANANSLAVFIKRCFTFMDRGVFKQINNYISCFAPGDPKTLPEYKEFLR 540
Qy 1071 VVCNHEHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHPLVGLLRREVGTALQEFR 1130
Db 541 VVCNHEHYIPLNLPMPFGKRIQRYODLQDYSLTDEFCRNHPLVGLLRREVGTALQEFR 600
Qy 1131 EYRLAISVLKXLLIKHSFDDRYASRHOARIATLYPLFGLLIENVORINRVDSPPV 1190
Db 601 EYRLAISVLKXLLIKHSFDDRYASRHOARIATLYPLFGLLIENVORINRVDSPPV 660
Qy 1191 NAGMTVKDESLALPAVNPPLVTPQKSTLDSNLHKDLGAI SGIASPYTTSTPNINSVNA 1250
Db 661 NAGMTVKDESLALPAVNPPLVTPQKSTLDSNLHKDLGAI SGIASPYTTSTPNINSVNA 720
Qy 1251 DSRGSLISTDSGNSLPERNSEKSNLSDKHQOQSTLGNSVVRCDKLDGSEIKSLMCFLYI 1310
Db 721 DSRGSLISTDSGNSLPERNSEKSNLSDKHQOQSTLGNSVVRCDKLDGSEIKSLMCFLYI 780
Qy 1311 LKSMDDALFTYWNKASTSEIMDFPTISEVCLHQPQYMKRYIARNOEGLAGIIVDRKSQ 1370
Db 781 LKSMDDALFTYWNKASTSEIMDFPTISEVCLHQPQYMKRYIARNOEGLAGIIVDRKSQ 833
Qy 1371 NSLTFPHSYGSHDADVLHOSLEANIATEVCLTALDTLSLFTLAFK 1416
Db 834 -GISVDNGYGSHDADVLHOSLEANIATEVCLTALDTLSLFTLAFK 878

```

RESULT 4

```

Q9B226 PRELIMINARY; PRT; 724 AA.
ID Q9B226;
AC Q9B226;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE BA15SN3.2.4 (KIAA1058) (Fragment).
GN BA15SN3.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

```

OX NCBI TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Smith M.;
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL61420; CAC27817.1; -
 FT NON TER
 SQ SEQUENCE 724 AA; 82617 MW; 7E2B345FE2FE63D6 CRC64;
 Query Match 33.8%; Score 3693; DB 4; Length 724;
 Best Local Similarity 97.1%; Pred. No. 3.3e-227;
 Matches 708; Conservative 1; Mismatches 6; Indels 14; Gaps 1;
 QY 531 TLEKDSAGNLDKNAKARSATYRQDSNKLSDNDMLKADLRKPKPKMAKLPEVILGNLDTITD 590
 DB 1 TLEKDSAGNLDKNAKARSATYRQDSNKLSDNDMLKADLRKPKPKMAKLPEVILGNLDTITD 60
 QY 591 NVSSDPNNVNSSYITPKQETCSKTPITEVEEVPVPCIPKTPQPTIYTNHLVYVPKYL 650
 DB 61 NVSSDPNNVNSSYITPKQETCSKTPITEVEEVPVPCIPKTPQPTIYTNHLVYVPKYL 120
 QY 651 KYDSOKSPAKARVIAICIEFKDSDEDSQPLKCIYGRPGGPVTRSAFAVILHHONPEF 710
 DB 121 KYDSOKSPAKARVIAICIEFKDSDEDSQPLKCIYGRPGGPVTRSAFAVILHHONPEF 180
 QY 711 YDEIKIELPTQHEKHLLLTFFHVSQDSSKSTKKRDVETQVGSMLPLKDKGRVVT 770
 DB 181 YDEIKIELPTQHEKHLLLTFFHVSQDSSKSTKKRDVETQVGSMLPLKDKGRVVT 240
 QY 771 SEGHIVSANTLPGYGYOELGNGRIYGPETKWDGSKPLKISTHVSSTVYTDQHLN 830
 DB 241 SEGHIVSANTLPGYGYOELGNGRIYGPETKWDGSKPLKISTHVSSTVYTDQHLN 300
 QY 831 FPOYCOKTSAGALNBLVKYKSHAMEGHVMAFLPTILNQLRVLTRATOESVAVN 890
 DB 301 FPOYCOKTSAGALNBLVKYKSHAMEGHVMAFLPTILNQLRVLTRATOESVAVN 360
 QY 891 VTRVLIHVVAQCHEGESHLRSYVYAYKABEYVASEKTVHEBLTKSMITLTKPSADF 950
 DB 361 VTRVLIHVVAQCHEGESHLRSYVYAYKABEYVASEKTVHEBLTKSMITLTKPSADF 420
 QY 951 LTNKLLKTSWFFDYVILKSMOHLIENSKVLLNQRPPASTYHAAVEVYVNMALPHITQ 1010
 DB 421 LTNKLLKTSWFFDYVILKSMOHLIENSKVLLNQRPPASTYHAAVEVYVNMALPHITQ 480
 QY 1011 KFRDNEASAKNAHSLAVFKRCFTMDRGFVKQNNNTYISGAPDPTLFEYKEEFLR 1070
 DB 481 KFRDNEASAKNAHSLAVFKRCFTMDRGFVKQNNNTYISGAPDPTLFEYKEEFLR 540
 QY 1071 VVCNHEHYIPLNLPMPFGKRIORVODLQDLSLTDFCRNHFVGLLREVGTAQOEFR 1130
 DB 541 VVCNHEHYIPLNLPMPFGKRIORVODLQDLSLTDFCRNHFVGLLREVGTAQOEFR 600
 QY 1131 EYRLIAISYVKNLLIHSFDDRYASRQARIATLYPLFGILLIENVORINRVDSPPV 1190
 DB 601 EYRLIAISYVKNLLIHSFDDRYASRQARIATLYPLFGILLIENVORINRVDSPPV 660
 QY 1191 NGMTYKDSIALPAVNPLVTPQKSTLNSLHKDLIGIASYTTSTPINSVRNA 1250
 DB 661 NGMTYKDSIALPAVNPLVTPQKSTLNSLHKDLIGIASYTTSTPINSVRNA 706
 QY 1251 DSRGSLIST 1259
 DB 707 PCSGGLST 715
 RESULT 5
 ID Q9VM04 PRELIMINARY; PRT; 1984 AA.
 AC Q9VM04;
 DT 01-MAY-2000 (TEMBLrel. 13, Created)
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TEMBLrel. 13, Last annotation update)

DE CG6630 protein.
 GN CG6630.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoekline R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Suton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriil J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burdle K.C., Busam D.A., Butler H., Cadieu B., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
 RA Jalali M., Kalush F., Karpn G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Meruliov G., Milghina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskehn D.R., Paclob J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter R., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasmann D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhou X., Zhu S., Zhu X., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Zheng L.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL; AB003617; AAF5824.1; -
 DR FlyBase; FBgn0031921; CG6630.
 DR InterPro; IPR001849; PH.
 DR Pfam; PF00169; PH; 1.
 DR SMART; SM00235; PH; 1.
 DR PROSITE; PSS0003; PH DOMAIN; 1.
 SQ SEQUENCE 1984 AA; 224875 MW; 5544A4A34022712F CRC64;
 Query Match 30.2%; Score 3305.5; DB 5; Length 1984;
 Best Local Similarity 37.0%; Pred. No. 9.6e-202;
 Matches 767; Conservative 343; Mismatches 615; Indels 347; Gaps 49;
 QY 45 KPKLIEPLDYENVYQKQTQILNDCLEMLFFPYDDFQTALIRQGRYICSTV----- 97
 DB 5 KPLVVEPIDFEFAFKAKNTVITQNDPQRRLILYPPDDVSEIIMPKRQRTNMAVSADRPDP 64
 QY 98 -----PAAEEEAQ----- 97
 DB 65 NEAIVCPHAGSPMITNGNGHGQVSRQSGISQNGSHNGNGNGHTSSSSSSSLNNSNGHQ 124
 QY 98 -----PAAEEEAQ----- 106
 DB 125 LSRKSSQCSNGSSSQKQKSSYSALSSITLRSNLAQPEVDEFAADGNADLVQGPQHSR 184

QY 107 ---SLPTECIKTNSDMHLVNYKXEDYSGEFRQJLBNKVVLDDKLPVHYVEDEEVDKDE 163
 DB 185 AECTFRQALVYTBPAKHLIHYKNAVAGNCHDLP5-161AEELVEEIEYIDADQDRID 243
 QY 164 DAASIGSQKGGITTKGMLYKGMNSA-----ISVTRSPFKRHPHLIQLQDGSYKREFLKD 219
 DB 244 EOMT-RSQADVTYTKQGYLLKGPDSASDRMPANIGKSPFRHRYCYLRQETDGYILIELHMD 302
 QY 220 -LOKEPKSIFLGLFYGV5FRNNKVRPAFELKM-ODKSSVLLADSEVEMEWITILNK 277
 DB 303 EKQGAKATIVMDFCTDV-VQNPKGGRFCFELRMTTGHSFTLAENEDPFQDWLSKISS 361
 QY 278 ILQNLNPEAMOKRND-----SHDEQSKLE-----GSGSLDYL-PELASA 322
 DB 362 VLAQN--RAOEKRVASLEHRPSIGSNP5PQLQPRAMPDPFTGTLKGLDQSLHPQLMKYG 419
 QY 323 REAEIKL--KSESVKLP--YLDPDAQULDFSSABEYK5PEEKGKRIYKCDLSFN 377
 DB 420 RETDSIALARREQRRLPACYQSPVK-----GSGSDNEQYREHFGRTLLTCHNLRER 474
 QY 378 LOCCVAENEB---GPTTVPEPFVTL5LPDIKYNRKISAD7VVDLNF5VRQMIATTS-P 433
 DB 475 LQ-CIPODESSAGVBOQVEPYITSLALPDAKANKLSEN7YFNNEQMAQDLPTPTVP 533
 QY 434 ALMNGSGPETGALRGILHEAMQY-----KQGISVYTCBP 471
 DB 534 SSVACQGVPRKGA-BGDERSTACQAPHSLPDGV5ABELRSNQOFQQLRGCLSLTAPHA 592
 QY 472 DIFVLARIEKVLQSGITTHCABPYMKS5DSSKYAQYLLKNAKQACQRLQGYRMPMAART 531
 DB 593 DYLVRLEKILQSGIAQVTEPYLLKAGKDPKLGQKYKAK5YAOHIGHYRPPFMAARP 652
 QY 532 LFKDASGNLD-----KNARSAIYRQDSNKL5NDMLKLABRKPKEKAKLVIIIGNDI 587
 DB 653 LFKQYSHEDVDPKKBEFSP5PIYQREPIKLDKEELIKLVDRKEBKSKLTIITIGSLKM 712
 QY 588 TIDNVSDEPNYNS5YIPTQFETCSKPTITEVEEFVPCIPKHTOPYTITNHLVYVP 647
 DB 723 QMGFLDQTTGCLSKSLAPLSTFSP5SKQSPTEVAE5FQ5E5BRDHPHTSPCNHLVYVP 772
 QY 648 KYLKIDSQSPAKANRIAICIEFKDSDEDSOPLKCIYRPGGPPVFTSAPAAVTLHHON 707
 DB 773 LSLQPSQCLF5RANITIVVELRDGD5Y5KPLKCIYRPGODLLV5QIACPVLHNVT 832
 QY 708 PEFYBEIKI5LPTOLHEKHLLTTPFV5CDNS5KSGSTKRD---VVENQVGY5MLPLK 764
 DB 833 PLYBEIKIKRLDGLFPEHLLP5FTV5CNLS-----KKRDANA5FEPPIGIALMLPLQ 887
 QY 765 DGRVVTSEQHIV5ANLPSGYLGYOELGMR--HYGPEIKWYDGGKPLKISTHLVSTVY 822
 DB 888 KNRICIEBQQLPVATLPPGYLSIQPLGK5GONGCPDICWIDNGRNLVTVGLARDSTYL 947
 QY 823 TODOH5LNFQYQKLT---ESGAQLGNELVYKLSLHPMEGHVIAFLPTILNQLFRV 878
 DB 948 TADOH5LNF5GCHERL5EGK5TGA5PAETETCKILKAAHIDMK5ILINYLPTVLELFTL 1007
 QY 879 LTRAQOE5AVVNTVYIIVVAQ5CHEEGLESH-L5SYVYAK5ABEYVASE---YKTV5E 934
 DB 1008 LVHTQ5EE5GLNVIRITINIH5L5DQAK5DLGL5YV5VFA-PY55Q5QARQARTVHG 1066
 QY 935 BLTK5MTTLKPS-ADFLTSNKLK5MFPFVLK5MAQHLIENS5VQLLNQ5RFP5SY 993
 DB 1067 ELCHRL5PYL5N5NP5FLV5N5P5R555IF5P5L5IK5M5QHL5L5AG5R5IM5AN5EP5K5Y 1126
 QY 994 HHA5VETV5M5LPHITQ5K5R5N5P5AS5KAN5SLA5VIFK5CTF5MD5RG5VFK5QINN5ISCF 1053
 DB 1127 GDRVEOL5IALMP5YITTR5ED55E5E5THL5NR5SLAK5RVQ5CL5YMR5GF5RLIR5CM5GF 1186
 QY 1054 ARGD5PTL5EYK5F5E5FL5V5CN5HEHY5PL5L5MP5G5GR5QR5QOD5LQ5D5L5DEC5R5NF 1113
 DB 1187 55GN5P5IL5EYK5F5N5L5Q5ICQ5EH5V5PL5L--P5VL5N5P5QR5P5M5QHL5L55Q5F5R5Q5F 1244
 QY 1114 LVGL5L5REV5G5ALQ5E5REV5L5A5SVL5KNL5IKH55F5DR5Y5AS5H5QAR5IAT5L5PL5GL 1173

DB 1245 L5GILLQELK5SLNEVGHVR5HALGIFKDL5LAK5H5LNR5YQ5KOL5HIAL5L5V5P5L5GV 1304
 QY 1174 IENQRI5VR5D5V5P5PV5NAG5TVKDE---SLAL5AV5N5PL5V5POK5STL5DN----- 1220
 DB 1305 MDN5HR5IDD5L5ESG5AT5P5NG5V5AD5AS5YTR5KLS5CSS5YV55K5OST5G5L5T5P5R5KR 1364
 QY 1221 -SLKH5DL5GAL5G5I5G5AP5YT5T5T-----P5INS5VR5N5D5NG5L5IST5D5G5NL 1265
 DB 1365 L5TH5M5DQ-----AS5P5R5IS5VH5KH5EN5Y5L5A5I5AG5P5IS5N5L5S5L5S-----N5D5SGH5Q 1414
 QY 1266 PE-----R5NSEK5NSL-DK5Q5S5TL5G5NV--R5CDL5D5E5IK5L5M5CL5Y5LK5MSD 1316
 DB 1445 D5TTIG5AT5ND5DT5VAL5NG5HR5SV5YH5AO5IL5RC5DF5S5VES5K5DL5L5GL5FL5I5KH5L5Q 1474
 QY 1317 DAL5FTY5N5K5AST5EL5M5D5FIT5E5V5C5LQ5FQ5Y5K5RYI-----AR5G5M5H5AR5L5Q 1365
 DB 1475 EOM5GM5W5Q5N5E5SET5LO5FL5L5D5L5C5LO5FR5Y5G5K5SV5IT5T5E5T5O5G5R5L5AK5N5TL5P5ART5P 1534
 QY 1366 L5G5L5DN-----SL5F5N5H5YGH--5P5AD5V5L5H5O---SL5E5N5I5ATE5V5CL5T5AL5D5TL5SL5FT 1412
 DB 1535 P5T5G5L5NS5Q5B5Q5P5SG5TL5N5Q5R5H5L5B5ID5TL5AN5Q5L5AL5E5SN5L5AT5E5GM5I5L5D5GL5YV 1594
 QY 1413 LAF5N5QL5I5AD5GH5N5PL5MK5VP5DY5L5C5PL5Q5H5Q5E5T5AL5KN5FT5AL5SL5YK5P5ST5Y5G5RA 1472
 DB 1595 LQ5FR-Q5LL5AD---5L5V5LP5K5AR5V5Y5L5F5L5Q5Q5S5R5L5SGH5V5PAL5RAP5IN5Y5A5VAL5FK5NA 1650
 QY 1473 DM5CAL5CE5IL5K5CN5K5L5S5I5R5E5AS5Q5L5Y5L5M5NN5P5Y5TK5S5F5V5R5H5L5O5V5I55V5Q5L 1532
 DB 1651 MLC5Q5M5Y5VEL5L5K5AC5D5SR5L5VE5IR5H5SC5AV5L5L5M5NS5E5F55GR5K5L5TR5H5L5Q5V5I55V5Q5M 1710
 QY 1533 ADV5G5IG5TR5FQ5Q5S5I5IN5C5AND5RL5IK5H5FS5SD5V5D5L5TK5I5RT5V5L5M5AT5Q5M5K5EH5ND 1592
 DB 1711 GNV5IG5L5N5N5AR5Q5B5G5I5S5I5NS5AND5DK5M5KG5F5M5E5V5D5L5RR5RT5V5L5M5A5TR5Q5AH5MD 1770
 QY 1593 PEM5L5VD5Q5Y5L5AK5Y5AST5PE5L5R5KT5W5L5D5MA5I5H5V5K5ND5L55A5M5CY5VH5T5AL5VA5E5Y5L5TRK 1652
 DB 1771 PER5L5LE5Q5Y5L5ANS5AST5PE5L5R5HT5W5L5Y5M5A5N5H5Q5N5L55A5C5H5L5I5AL5M5CE5YL--- 1827
 QY 1653 EAV5Q5M5E5P5L5L5H5S5AC5LR5R5RG-V5P5R5Q5CT5AR5V5T5P5ND-BE5AS5M5E5V5Q5M5DV5FN 1710
 DB 1828 -----R5K5G5CT5L5WS5TA5G5K5I5ST5N5PL5B5Q5G5L5K5D5AG5Q5D5O5YT 1868
 QY 1711 EDV5L5M5L5E5Q5AD5GL5M5K5ER5Y5EL5ADI5K5L5I5PI5E5K5RD5E5RL5A5H5L5D5TL5RA5SK5Y5E 1770
 DB 1869 EOM5L5E5Q5L5K5AD5PL5DR5ER5E5C5G5E5L5Y5L5L5P5Y5E5R5S5Q5DL5A5H5C5E5H5L5Q5A5NK5-- 1926
 QY 1771 VMS5GR5LL5G5TY5F5VA5F5G5QA5Q5Y5Q5FT5D5E5T5D5E5G5F5E5D5E5G5E5Y5K5E5P5L5L5S5E5I5Q 1830
 DB 1927 -----V5M5Y5F5E5D5H5A5I5E5V5Y5K5E5P5L5T5L5S5E5I5S5E 1954
 QY 1831 RLL5K5Y5SD5K5F5G55EN5Y5N5I5Q5D5G5K5N5PK5D5L5DK 1862
 DB 1955 RLAK5Q5Y5K5E5F5G5AD5V5K5M5IDS---5P5V5L5DK 1983
 QY 6 Q63603
 ID Q63603 PRELIMINARY; PRT; 738 AA.
 AC Q63603;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAY-2000 (Tremblrel. 13, Last annotation update)
 DE TRG protein (Fragment).
 GN TRG.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN NCB1_TaxId=10116;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FISCHER;
 RA Pianeze L., Porcellini A., Avvedimento V.E., D'Esposito F.,

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC009134; AA009134.1; -
SQ SEQUENCE 417 AA; 48431 MW; 6FCCEC3A90E059678 CRC64;

Query Match 19.1%; Score 2094; DB 11; Length 417;
Best Local Similarity 94.4%; Pred. No. 1,9e-125;
Matches 407; Conservative 6; Mismatches 4; Indels 14; Gaps 1;

QY 1677 VFRQGTAFRVITPNIDEASMMEDVGMQDVHFNEDVLMELBOCADGLMKARVELIND 1736
DB 1 MFRQGTAFRVITPNIDEASMMEDVGMQDVHFNEDVLMELBOCADGLMKARVELIND 60
QY 1737 IYKLIPIYKRRDFERLAHLYDTLHRAYSKTEVMSGRRLGTYFRVAFGQAQYOF 1796
DB 61 IYKLIPIYKRRDFERLAHLYDTLHRAYSKTEVMSGRRLGTYFRVAFGQ----- 114
QY 1797 TDBETDEGFEEDGKEYIYKPKLTPLEISQRLKLYSKFGSEBNMAMIDSGKNP 1856
DB 115 -----GFEEDGKEYIYKPKLTPLEISQRLKLYSKFGSEBNMAMIDSGKNP 166
QY 1857 KDLDSKAYIQVHVHPFPEKELOERKTEFERSHNIIRPFMPPTONGKRGQVEBOC 1916
DB 167 KDLDSKAYIQVHVHPFPEKELOERKTEFERSHNIIRPFMPPTONGKRGQVEBOC 226
QY 1917 KRRTILTAIHCFPYVKRIPVMYQHTDLPFEVAIDEMSKYAEELROLCSNAEVMIKL 1976
DB 227 KRRTILTAIHCFPYVKRIPVMYQHTDLPFEVAIDEMSKYAEELROLCSNAEVMIKL 286
QY 1977 QLLQSSVSQVNAAGPLAYARAFLDNTNTRYPDNRYKLYKXFRQPEACGALAVNER 2036
DB 287 QLLQSSVSQVNAAGPLAYARAFLDNTNTRYPDNRYKLYKXFRQPEACGALAVNER 346
QY 2037 LIKEDOLEJOEEMKANYREMAKELSEIMHOICPLEKTSVLPNSLHIFNAISGTPTSM 2096
DB 347 LIKEDOLEJOEEMKANYREMAKELSDIMHOICPLEKTSVLPNSLHIFNAISGTPTSTV 406
QY 2097 VHGMTSSSVV 2107
DB 407 VGLTSSSVV 417

RESULT 11
ID Q9GM45 PRELIMINARY; PRT; 425 AA.
AC Q9GM45;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (T-EMBLrel. 16, Last annotation update)
DE Hypothetical 48.6 kDa protein (Fragment)
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN PARIETAL LOBE;
RA Oeada N., Hida M., Kuenda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RT Suzuki Y., Sugeno S., Hashimoto K.;
RT Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB049841; BAB16727.1; -
KW Hypothetical protein
FT NON_TER 425 425
SQ SEQUENCE 425 AA; 48574 MW; E05B25BEC9984003 CRC64;

Query Match 19.1%; Score 2085.5; DB 6; Length 425;
Best Local Similarity 93.8%; Pred. No. 6.7e-125;
Matches 408; Conservative 2; Mismatches 2; Indels 23; Gaps 1;

QY 1358 MHAARIQQLGSLDNLTFNHSYGHSDADVLHOSLLEANIATEVCITALDTLSFTLAFKN 1417
DB 1 MHAARIQQLGSLDNLTFNHSYGHSDADVLHOSLLEANIATEVCITALDTLSFTLAFKN 60
QY 1418 QLLADHGNPLMKKYPDYVLCFLOHROSETLAKVFTLRSLIYKPFSTFGGRADMCQA 1477
DB 61 QLLADHGNPLMKKYPDYVLCFLOHROSETLAKVFTLRSLIYKPFSTFGGRADMCQA 120
QY 1478 LCYEILKCNLSLSSIRTEASQLLYFLMRNNDYTGKSPYTHQLQVLIISQLIADYVG 1537
DB 121 LCYEILKCNLSLSSIRTEASQLLYFLMRNNDYTGKSPYTHQLQVLIISQLIADYVG 180
QY 1538 ICGTRFOOSLTIINNCANSRLIKTSPSDYKDLTKIRTYLMTAQMKEHNDPENLV 1597
DB 181 ICGTRFOOSLTIINNCANSRLIKTSPSDYKDLTKIRTYLMTAQMKEHNDPENLV 240
QY 1598 DLQYSLASVASTPELRKTWLDMSARHVKNGDSEAMCYHVTAALAEVLTREAVQM 1657
DB 241 DLQYSLASVASTPELRKTWLDMSARHVKNGDSEAMCYHVTAALAEVLTREAVQM 295
QY 1658 BPPLPHSHSACLRSGVFRQGTAFRVITPNIDEASMMEDVGMQDVHFNEDVLMEL 1717
DB 296 -----GMFRQGTAFRVITPNIDEASMMEDVGMQDVHFNEDVLMEL 337
QY 1718 LEQCADGLMKARVELIADLYKLIPIYKRRDFERLAHLYDTLHRAYSKTEVMSGRRL 1777
DB 338 LEQCADGLMKARVELIADLYKLIPIYKRRDFERLAHLYDTLHRAYSKTEVMSGRRL 397
QY 1778 LIGTYFRVAFGQA 1792
DB 398 LIGTYFRVAFGQVS 412

RESULT 12
ID Q9BMX9 PRELIMINARY; PRT; 696 AA.
AC Q9BMX9;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE B125M24.1 (KIAA1058 protein) (Fragment).
GN B125M24.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilson S.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL391237; CAC36036.1; -
FT NON_TER 696 696
FT NON_TER 696 696
SQ SEQUENCE 696 AA; 81097 MW; ASBCBF055A94F7A8 CRC64;

Query Match 18.4%; Score 2007; DB 4; Length 696;
Best Local Similarity 53.0%; Pred. No. 1.5e-119;
Matches 390; Conservative 122; Mismatches 164; Indels 60; Gaps 10;

QY 600 VNSSTIPKQRE-TCSKTPTTFEVEBEFVPCIPKXHTOPTITTNHLYVPKYLKDSQSF 658
DB 2 ITTSVPLKPEKQCN--ITVEVEBEFVEMTKCYPTIYKHLHYVPLDKYDSQKTF 59
QY 659 AKARIIATCIFKDSDEDSQPLKCIYGRPGAPVFTSRAPAAVYHHHONPEFYDEIKTEL 718
DB 60 AKARIIATCVERDSDESDALKCIYGRPGASFTTNAIVVSHHONPEFYDEIKTEL 119
QY 719 PTOLHEKTHLLTFPHVSCDSSKSTKRDVETQVGSWLPPLKDGKVVTSQHIPIVS 778

```
Db 120 PHLHDKHLLFTFYHSCBINTKGTQDVEIPVGFANVPLKDGRIITPEOOLPVS 179
Qy 779 ANLPSCYLGLYOEIGMGRHNGPEIKKVDGKPLKSTHLSVFTYVODOLHNFPOYCOCKT 838
Db 180 ANLPSCYLGLYND-----QCNVDIKWYDGNKAPLKLSHLESTIYTODLHVHFHHCQL 234
Qy 839 ESGAOLAGNELVKYLSLHMEGHVNIAPLPTLNLQFVLTAAQOE-EVAVVTRVLIH 897
Db 235 QGSGKEVPGLLKYLKCLHAMEIQWIGLPLVIMQLPVLNTMTHEDDVPINCTVLIH 294
Qy 898 VVAQCHBEGSLHSYKAYKAEPYVASSEKTYHEELTKMTTLTKSADPLSNKLL 957
Db 295 IYKCHBEGSLDYLSPFIKYSFPEKPSAPOQLIHETATMTIALIKOSADPLSNKLL 354
Qy 958 KTSWPEFDVTKSMOHLIENSKVKLRRORFPASVHNAVETVNMMLHITQKPRDNE 1017
Db 355 KTSWPEFDVTKSMATYLLBENKIKLPKQRFPEYTHVHLSLLAIIEHVITRIAEIPD 414
Qy 1018 ASKNANHSIAVPIKRCFTFMDRGFVFKQINNYISCPAPDPKTLFEYKEFPLRVVGNH 1077
Db 415 ESRNNVYSLASFILKCLTLMDFGFIENLNDYISGSPKDPKYLAKYKEFLQTIQNH 474
Qy 1078 YPLPLPMPFGKRIQRYODLDYSLTDFPCRNHFLVGLLREVGTALQEFREVLI 1137
Db 475 YPLPLPMPAFKPKIQRYODSNLEYSLSDEYCHHFLVGLLRETSIALQDVEIRYTAI 534
Qy 1138 SYLKNLLIHSFDDRYASRSHARITLYLPLFGLIENVORINVDVSPFPNMGMTYK 1197
Db 535 SYLKNLLIHAAPDTRIOHKOQAKIAQLYPLVGLLENIQRIAGD----- 581
Qy 1198 DESLALPAVNPVTPQKSGSTLNSL-----HKDLGAIISGLASPYTSTPINSVRND 1251
Db 582 -----LYSCAMPNSLMLGFRMOMELREKIQEVS-----SQKEQDDQ 621
Qy 1252 SSGSLIST-DSGNSLPERNSEKSNL-----DKHQSGSLGNSVVRCDKLDQSI 1300
Db 622 IRATLVKIDYFKILPLKMAVENETQYSWLYLCEBDSKTRQST-RSSVSGVRLDQYEI 680
Qy 1301 KSLMCELYTKMSMD 1316
Db 681 KSLMCELYTKMISE 696
```

RESULT 13

```
06XK11 PRELIMINARY; PRT; 534 AA.
ID 06XK11, PRELIMINARY; PRT; 534 AA.
AC 06XK11, PRELIMINARY; PRT; 534 AA.
DT 01-MAR-2002 (TRENBLrel. 20, Created)
DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE BB12804.1.2 (Novel protein, isoform 2) (Fragment).
GN BB12804.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lawlor S.
RA Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL, AL391280; CDD21443.1; -.
FT NON_TER 1 534
FT TER 534
SQ SEQUENCE 534 AA; 61426 MW; 8A0PD929A5DD41F CRC64;
Query Match 18.2%; Score 1985.5; DB 4; Length 534;
Best Local Similarity 63.3%; Pred. No. 2, 4e-118;
Matches 303; Conservative 76; Mismatches 71; Indels 75; Gaps 2;
```

```
Qy 1463 PSTFEGADNCAALCYELKCKNKLSSIRTSQQLYFLMRNNPDYTGKSPVTRTL 1522
Db 1 PFSAPFKGHVNMCAACYEVLKCKTSKISTREASALLYLMLRNNEFYTKRTPLRTHL 60
```

```
Qy 1523 QVITISVQLIADVVGIGSTRFQOQSLIINNCAUSDRLIKHTSPSSDVNLTIRIRTVLMA 1582
Db 61 QITIAVSQILIDVALSGSGRFOQESFLIINNPNANSRPMKATAPRAEVDLTIRIRTVLMA 120
Qy 1583 TPAQKEHNDPMLVDLOYSLAKSYASTPELRYKTLWDSMARIIYKGLSBAACVYAVT 1642
Db 121 TPAQKEHNDPMLVDLOYSLAKSYASTPELRYKTLWDSMARIIYKGLSBAACVYAVT 180
Qy 1643 ALVAEYLRKEAVNQWEPFLPHSHSACLRASRGVFRQGCCTAPRYITPDIIBASMDV 1702
Db 181 ALVAEYLRKKE-----LFNGCSAFKIIYNIDBEGAMKEDA 217
Qy 1703 GQWDVHFNEVDYLMELBEOCADGLMKAERYELIADYKLIPIYKRRDPERRLAHLYDTLH 1762
Db 218 GQWDVHFNEVDYLMELBEOCADGLMKAERYELIADYKLIPIYKRRDPERRLAHLYDTLH 266
Qy 1763 RAYSKVTEVMSGRRLGTYRRVAFPGQAQYQFTDSETDVGEFPEDEGKEYIYKEPL 1822
Db 267 -----ESFPEREDGKEYIYKEPL 285
Qy 1823 TPLSEISQRLKLYSDEKGSENVKMIODSGVNPDKLDSKAYIQTHTVPEDEKELOE 1882
Db 286 TGLSEISQRLKLYSDEKGSENVKMIODSGVNPDKLDSKAYIQTHTVPEDEKELOE 345
Qy 1883 RKTERRSHNIRRFEMFEMFTQTRGROGVEBQCKRRTILTAIHCFPYVKRIPIWYOH 1942
Db 346 RKTERRSHNIRRFEMFEMFTQTRGROGVEBQCKRRTILTAIHCFPYVKRIPIWYOH 405
Qy 1943 TDLANPIEVALIDMSKRVABLQCSSAEVDMIKQLKQSSVSVOVNAAPLAPAPLDD 2002
Db 406 INLKPIDVATBEIKOKTAELOKLCSTVDVMIQLOLKGCVSVQVNAAPLAPAPLDD 465
Qy 2003 TMTKVPYNNKYKLLKEVROPEVACGALANVERIKEDOLYEOSEMANTREMAKELSE 2062
Db 466 SOASTYPPKVSSELMKMRFTIQACSIABEINERIKEDQVYHGLKSNFRDVKELSD 525
Qy 2063 IMHEQ 2067
Db 526 IMHEQ 530
```

RESULT 14

```
045377 PRELIMINARY; PRT; 2374 AA.
ID 045377, PRELIMINARY; PRT; 2374 AA.
AC 045377, PRELIMINARY; PRT; 2374 AA.
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE F22612.5 protein.
GN F22612.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Eubaditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Leonard N.
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Almscough R., Anderson K., Baynes C., Berke N.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kerehaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,
RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Watson R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
```



```
Qy 1642 TALVAEYLTRKAVOWEPPLPHSHSACLRSRGVPROGCTAFRVITPNIDEASMMED 1701
Db 1108 AALVAEYLSWLE-----DHSY-----LPVGSVSFQNISSNVLEESVYSED 1147
Qy 1702 VGMQD-----VHFNEDVLMELLEQCADGLMKAERYELIADYKLIIPYEKRDPERL 1754
Db 1148 TUSPDEBGVCAGQYFTESGLVGLLEQAAELFSTGLEYETVNEVYKLVIPLEAHRERFKL 1207
Qy 1755 AHLVDTLHRAYSKVTEVMSGRRLGTYPVAFPGQAQYQFTDSETDVSGFEDEDGKE 1814
Db 1208 TLTHSKLQRAFDIYNKDH--KMFGTFRVGFSGK-----FGDLDEQE 1250
Qy 1815 YIYKEPKLTPLSEISQRLKLYSDKFSENVKMIQDSGKVNPKDLDSKYAYIQTVHVPF 1874
Db 1251 FVYKEPATIKLPEISHLEAFYGCFCGAEFEVIXDSTPVDKTKLDPNKAYIQITFVEPY 1310
Qy 1875 FDEKELOERKTEPERSHNIRRFMEPFTOTGKQGVBEQCKRTILTAHCFPYVKKR 1934
Db 1311 FDEYEMKDRVYFEKJFNLRFFMYTTPFTLEGRPRGELHEOYRRNTVLITMHAFPYIKTR 1370
Qy 1935 IPVMYQHHTDNLNIEVAIDEMSKVAELRQLCSSAEVDMIKLOKLQGSVSQVNAEPLA 1994
Db 1371 ISVIQKEBEVLTPIEVAIEDMKKKTQLAVAINQBPDPANMLQVLOGSVGATVNOGPLE 1430
Qy 1995 YARAFLLD--TNTKRY-PDNKVKLLKEVPROFVEACGOALAVNERLIKEDOLEYQEMKA 2051
Db 1431 VAQVFLAEIPADPKLYRHHNKLRL---CFKEFIMRCGEAENKRLITADQREYQOELKK 1487
Qy 2052 NYREMAKELSEIMHEOI 2068
Db 1488 NYNKLKENLRPMIERKI 1504
```

Search completed: July 14, 2003, 18:21:10
Job time : 106.749 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:16 ; Search time 26.0723 Seconds
(without alignment)
3843.329 Million cell updates/sec

Title: US-09-815-379-6

Perfect score: 3898
Sequence: 1 MRLGAAVAALGRGRAPAS.....INKYLEATGQLPVKKKAKN 752

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 13350620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: A.Geneseq.101002.*
2: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT.*
3: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT.*
4: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT.*
5: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT.*
6: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT.*
7: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT.*
8: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT.*
9: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT.*
10: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT.*
11: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT.*
12: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT.*
13: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT.*
14: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT.*
15: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT.*
16: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT.*
17: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT.*
18: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT.*
19: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT.*
20: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT.*
21: /SID52/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT.*
22: /SID52/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT.*
23: /SID52/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT.*
24: /SID52/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3898	100.0	752	AAE11888	Angiogenesis assoc
2	3840.5	98.5	751	AAE11888	Human GTP-binding
3	2521.5	64.7	729	ABE60766	Drosophila melanog
4	2275.5	58.4	442	AAV01167	Polypeptide fragme
5	2113.5	54.2	409	ABE89506	Human polypeptide
6	1489	38.2	692	AAU35975	Helicobacter pylori
7	1454	37.3	700	AAU35467	Haemophilus influe
8	1431	36.7	741	AAU40036	Propionibacterium
9	1427	36.6	704	AAE15891	E. coli proliferat
10	1427	36.6	704	AAU34770	E. coli cellular p

11	1412	36.2	705	22	AAE90301
12	1406	36.1	702	22	AAU36265
13	1387	35.6	695	23	ABE48403
14	1374	35.2	704	22	AAU38372
15	1374	35.2	692	23	ABP30759
16	1352.5	34.7	691	22	AAU33481
17	1352.5	34.7	693	19	AAU80722
18	1352.5	34.7	693	22	AAU34888
19	1352.5	34.7	693	22	AAU37640
20	1352.5	34.7	693	22	AAU01093
21	1351.5	34.7	715	22	AAU36850
22	1351.5	34.7	715	22	AAU37143
23	1346.5	34.5	692	23	ABP27487
24	1338	34.3	693	22	AAU33742
25	1318	33.8	705	23	ABP40511
26	1315	33.7	702	20	AAU37794
27	1308	33.6	709	23	ABE56515
28	1178.5	30.2	651	20	AAU35171
29	1175	30.1	245	22	AAU20463
30	1107	28.4	223	21	AAE42032
31	944	24.2	191	21	AAE42257
32	896	23.0	207	22	AAU20620
33	784.5	20.1	1093	22	ABE62821
34	772	19.8	719	22	AAU38875
35	692.5	17.8	735	22	AAE96105
36	686	17.6	332	18	AAE20914
37	568.5	14.6	641	13	AAE30089
38	541.5	13.9	842	21	AAU3637
39	541.5	13.9	842	21	AAU3630
40	539	13.8	639	23	ABP30306
41	539	13.8	644	23	ABP27486
42	538.5	13.8	514	23	ABP28786
43	535.5	13.7	514	23	ABP30078
44	530	13.6	639	9	AAE80147
45	529.5	13.6	523	23	ABE53656

ALIGNMENTS

RESULT 1	AAE11888	AAE11888 standard; Protein; 752 AA.
ID	AAE11888	
AC	AAE11888;	
XX		
DT	18-DEC-2001 (first entry)	
XX		
DE	Angiogenesis associated hmt-elongation factor G (hEF-G) protein.	
XX		
KW	Angiogenesis associated protein; AAF; cytosolic; cardiac; gene therapy; ophthalmological; vulnary; myocardial infarction; macular degeneration; diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine; rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation; cancer; therapeutic; diagnostic; hmt-elongation factor G; hEF-G.	
OS	Unidentified.	
XX		
PN	WO200170808-A2.	
XX		
PD	27-SEP-2001.	
XX		
PF	22-MAR-2001; 2001WO-US09609.	
XX		
FR	22-MAR-2000; 2000US-191134P.	
XX		
PA	(CURA-) CURAGEN CORP.	
XX	(GETH) GENENTECH INC.	
XX		
PI	Rastelli LK, Gerritsen M;	
XX		
DR	WPI, 2001-602775/68.	
XX		
DR	N-PSDB; AAD19117.	

C glutamicum prote
Pseudomonas aerugi
Listeria monocytog
Salmonella typhi c
Streptococcus poly
Enterococcus faeca
S. pneumoniae elon
Enterococcus faeca
Streptococcus pneu
CPR 96 protein seq
Staphylococcus aur
Staphylococcus aur
Streptococcus poly
Staphylococcus aur
Staphylococcus epi
Protein involved i
Lactococcus lactis
Lactococcus lactis
Protein involved i
Human secreted pro
Human ORFX ORP1796
Human ORFX ORP2021
Human secreted pro
Drosophila melanog
Human BRCA-1 tumou
Futative P. abyssi
E. pylori cytoplas
Ter gene product d
Amino acid sequen
Lactococcus lactis
Streptococcus poly
Streptococcus poly
Streptococcus poly
Tetracycline resis
Lactococcus lactis

XX Novel angiogenesis associated polypeptides and polynucleotides encoding
PT the polypeptides, useful for modulating angiogenesis and for treating
PT tumors and cancers -
XX
PS Claim 1; Page 15-18; 159pp; English.
XX
CC The invention relates to angiogenesis associated proteins (AAP) and their
CC corresponding cDNA molecules, which are useful for modulating
CC angiogenesis. AAP proteins and nucleic acids are useful for promoting
CC wound healing, for example after organ transplantation, and in the
CC treatment of tumours, myocardial infarction, cancers, diabetic
CC retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.
CC AAP proteins and DNA's are useful in potential prophylactic and
CC therapeutic applications implicated in a variety of disorders including
CC those related to angiogenesis, and also in diagnostic applications.
CC AAP cDNA is also useful in gene therapy. The invention also relates to
CC a method for screening a tissue sample for tumourigenic potential. AAP
CC proteins are used to screen drugs or compounds that modulate AAP activity
CC or expression as well as treating disorders characterised by insufficient
CC or excessive production of AAP or production of AAP forms that have
CC decreased or aberrant activity compared to the wild type protein, or
CC modulate biological function that involve AAP. The present sequence
CC is hmt-elongation factor G (hmt-G) protein which is an angiogenesis
CC associated protein (AAP) of the invention. Hmt-EF-G is strongly induced
CC in an in vitro model of angiogenesis and is also used as an attractive
CC therapeutic target to treat disease states associated with mitochondrial
CC dysfunction.
XX
SQ Sequence 752 AA;
Query Match 100.0%; Score 3898; DB 22; Length 752;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 752; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRLGAAVAALGRGAPASLGWQRKQVNMKACRWSSSGVTPNEKIRNIGISAHDSGKT 60
DB 1 MRLGAAVAALGRGAPASLGWQRKQVNMKACRWSSSGVTPNEKIRNIGISAHDSGKT 60
QY 61 TLTERVLYYTGRTAKHNEVKGKGVAVNDMSLEORGTIOSATFTMKQVNNITD 120
DB 61 TLTERVLYYTGRTAKHNEVKGKGVAVNDMSLEORGTIOSATFTMKQVNNITD 120
QY 121 TPGHVDTIEVERALRYLDGAVLVCAGVGQCOTVTVMRQMRVYVPELTFINKLDRMG 180
DB 121 TPGHVDTIEVERALRYLDGAVLVCAGVGQCOTVTVMRQMRVYVPELTFINKLDRMG 180
QY 121 TPGHVDTIEVERALRYLDGAVLVCAGVGQCOTVTVMRQMRVYVPELTFINKLDRMG 180
DB 121 TPGHVDTIEVERALRYLDGAVLVCAGVGQCOTVTVMRQMRVYVPELTFINKLDRMG 180
QY 181 SNPARALQOMRSKLNNTAFMOJPMGLEGNFKGIIVDLIERRAYFPDGSQIVRYGRI 240
DB 181 SNPARALQOMRSKLNNTAFMOJPMGLEGNFKGIIVDLIERRAYFPDGSQIVRYGRI 240
QY 241 ELBAATDHRQELIECVANSDEOLGEMFLEKIPISIDKLAIIRBATLKRSFPPVLSA 300
DB 241 ELBAATDHRQELIECVANSDEOLGEMFLEKIPISIDKLAIIRBATLKRSFPPVLSA 300
QY 301 LKNKGVOPLDAVLEVLNPNSEVQNTAAILNKDSDSEKTKILNNSRHNHPVGLAFPL 360
DB 301 LKNKGVOPLDAVLEVLNPNSEVQNTAAILNKDSDSEKTKILNNSRHNHPVGLAFPL 360
QY 361 EVGRFGOLTYVRSYOGELKKGGDTTYNTRTKRQLORLARMDAMEASTEEVYADICA 420
DB 361 EVGRFGOLTYVRSYOGELKKGGDTTYNTRTKRQLORLARMDAMEASTEEVYADICA 420
QY 421 LFGIDCASGDTFTDKANSGLSMESIHVPDVIISANKPSKNDLEKPSKGIGFTEDPT 480
DB 421 LFGIDCASGDTFTDKANSGLSMESIHVPDVIISANKPSKNDLEKPSKGIGFTEDPT 480
QY 481 FKYPPTENKETVYISGGELEHLEIYAQRLEREYGCCTGPKVARETITAPVPDFTH 540
DB 481 FKYPPTENKETVYISGGELEHLEIYAQRLEREYGCCTGPKVARETITAPVPDFTH 540
QY 541 KKGSGAGQYGVKIVLEPLDPEDYTKLEFSDTFPSNIPKQFPAVEKGLDACEKGPL 600

DB 541 KKGSGAGQYGVKIVLEPLDPEDYTKLEFSDTFPSNIPKQFPAVEKGLDACEKGPL 600
QY 601 SGHKLSGLRFVLDQDAHMHVDSNEISPIRAGGALKOLANATLILPIMAVEVPANE 660
DB 601 SGHKLSGLRFVLDQDAHMHVDSNEISPIRAGGALKOLANATLILPIMAVEVPANE 660
QY 661 FCGQVIAGINRRHGVITGQDGVEDFTLYADVPPLNDMGYSTELASCTEGKGYTMEYSR 720
DB 661 FCGQVIAGINRRHGVITGQDGVEDFTLYADVPPLNDMGYSTELASCTEGKGYTMEYSR 720
QY 721 YQPCLPSTQEDVYINKYLEATGQLPVKKGAKN 752
DB 721 YQPCLPSTQEDVYINKYLEATGQLPVKKGAKN 752
RESULT 2
ID AAB68560 standard; Protein; 751 AA.
AC AAB68560;
XX
DT 19-APR-2001 (first entry)
XX
DE Human GTP-binding associated protein #60.
XX
DE Human; guanosine triphosphate binding associated protein; GTP; GBAF;
XX inflammation; AIDS; Addison's disease; anaemia; arteriosclerosis; asthma;
XX autoimmune disorder; hepatitis; multiple sclerosis; cancer; diabetes;
XX osteoporosis; psoriasis.
XX
OS Homo sapiens.
XX
PN WO200105970-A2.
XX
PD 25-JAN-2001.
XX
PF 19-JUL-2000; 2000WO-US19698.
XX
PR 19-JUL-1999; 99US-0144595.
XX 23-AUG-1999; 99US-0150460.
XX 15-OCT-1999; 99US-0159849.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Tang YT, Bandman O, Hillman JL, Lai P, Au-Young J;
PI Reddy R, Yang J, Baughn MR, Lu DM, Azimzal Y, Patterson C;
XX WPI; 2001-091972/10.
DR N-PSDB; AAF58360.
XX
PT New guanosine triphosphate-binding associated protein (GBAF) and their
PT encoding nucleic acids, useful for treating and/or diagnosing diseases
PT associated with GBAF expression, such as cancer, diabetes and asthma -
XX
PS Claim 1; Pages 178-180; 233pp; English.
XX
XX The present invention relates to novel human guanosine triphosphate
XX (GTP)-binding associated proteins (GBAFs; AAB68501-AAB68566) and their
XX coding sequences (AAF58301-AAF58366). The proteins and coding sequences
XX of the present invention are useful for treating a variety of disorders
XX including inflammation, AIDS, Addison's disease, anaemia,
XX arteriosclerosis, asthma, autoimmune disorders, Grave's disease,
XX hepatitis, multiple sclerosis, cancer, diabetes, osteoporosis and
XX psoriasis.
SQ Sequence 751 AA;
Query Match 98.5%; Score 3840.5; DB 22; Length 751;
Best Local Similarity 98.7%; Pred. No. 0;
Matches 742; Conservative 3; Mismatches 6; Indels 1; Gaps 1;
QY 1 MRLGAAVAALGRGAPASLGWQRKQVNMKACRWSSSGVTPNEKIRNIGISAHDSGKT 60

QY 644 LCLLEPIMAVEVAPNEFOGVIAGINRRHGVTITGDDVEDYFTLYADVPNDMFYSTTE 703
 DB 621 WQLEBPIMLVEVAPNEFOGVIAGINRRHGVTITGDDVEDYFTLYADVPNDMFYSTTE 680
 QY 704 LRSECTGKSGEYTMESRYOCPLESTOEDVINKYLEATGQLPVKKGAKN 752
 DB 681 LRSTGQKSGEYTMESRYOCPLESTOEDVINKYLEATGQLPVKKGAKN 729

RESULT 4
 AAY01167
 ID AAY01167 standard; Protein; 442 AA.
 AC AAY01167;
 XX
 XX 18-MAY-1999 (first entry)
 DT
 XX
 DE Polypeptide fragment encoded by gene 10.
 XX
 XX Human; secreted protein; gene therapy; protein therapy; tissue; cancer;
 KM tumour; neurodegenerative disorder; leukaemia; autoimmune disease; AIDS;
 KM developmental abnormality; foetal deficiency; Alzheimer's disease;
 KM cognitive disorder; schizophrenia; immunological disorder; mood disorder;
 KM immune deficiency disease; respiratory disorder; arthritis; skeletal;
 KM haematopoietic disorder; neural; osteoporosis; metabolic disorders;
 KM cardiovascular; endocrine; gastrointestinal; asthma; diagnosis.
 XX
 OS Homo sapiens.
 XX
 PN MO9901020-A2.
 PD 14-JAN-1999.
 XX
 PF 30-JUN-1998; 98MO-US13608.
 XX
 PR 12-SEP-1997; 97US-0058663.
 PR 01-JUL-1997; 97US-0051381.
 PR 01-JUL-1997; 97US-0051480.
 PR 12-SEP-1997; 97US-0058598.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Carter KC, Endress GA, Peng P, Rosen CA, Ruben SM;
 DR N-PSDB; AAX22120.
 XX
 XX MPI; 1999-105683/09.
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, immune deficiency diseases or blood
 PT disorders
 PS
 XX
 PS Disclosure; Page 17; 17pp; English.
 XX
 XX The invention relates to nucleic acid sequences (AAX22111 to AAX22134)
 CC encoding human secreted proteins (AAY01135 to AAY01158). The secreted
 CC protein gene sequences are deposited with the ATCC under deposit number
 CC ATCC 209118. Host cells comprising recombinant vectors containing the
 CC nucleic acid sequences are used for the recombinant production of the
 CC secreted proteins. The polynucleotide and amino acid sequences are useful
 CC for are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Pathological conditions can
 CC be also diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new
 CC polynucleotides. Specific uses are described for each of the
 CC polynucleotides, based on which tissues they are most highly expressed
 CC in, and include developing products for the diagnosis or treatment of
 CC cancer, tumours, developmental abnormalities and foetal deficiencies,
 CC autoimmune diseases, lymphomas, Alzheimer's and cognitive disorders,
 CC schizophrenia, immunological disorders, immune deficiency diseases
 CC (AIDS), mood disorders, respiratory disorders, arthritis, asthma,
 CC haematopoietic disorders, neural disorders, skeletal disorders,
 CC osteoporosis, metabolic disorders, cardiovascular disorders, endocrine

CC disorders or gastrointestinal disorders. The polypeptides are also useful
 CC for identifying their binding partners. The present sequence represents a
 CC polypeptide fragment encoded by a gene of the invention (see descriptor
 CC line for gene number).
 XX
 SQ Sequence 442 AA;
 Query Match 58.4%; Score 2275.5; DB 20; Length 442;
 Best Local Similarity 98.6%; Pred. No. 5.4e-186;
 Matches 437; Conservative 1; Mismatches 4; Indels 1; Gaps 1;

QY 310 LDAVLEIYPNPEVONVAILNKKDSSKEKTKLNMSSRNHSPFGIAFPLEVGFRGQLT 369
 DB 1 LDAVLEIYPNPEVONVAILNKKDSSKEKTKLNMSSRNHSPFGIAFPLEVGFRGQLT 60
 QY 370 YVRSYQGBLKKDDITYNTRTRKKVRLQRLARHADMEASTEEVYAGDICALFGIDCASG 429
 DB 61 YVRSYQGBLKKDDITYNTRTRKKVRLQRLARHADMEASTEEVYAGDICALFGIDCASG 119
 QY 430 DFTDKANSGLSMESIHVPDPVSIAMKPSNKNDEKFSKIGRTREDPPTFKVYFDTEN 489
 DB 120 DFTDKANSGLSMESIHVPDPVSIAMKPSNKNDEKFSKIGRTREDPPTFKVYFDTEN 179
 QY 490 KETVISGMBLHLEIYAQRLEREYCCPCITGKPKYAFRETTIAPVPFPDTHKKQSGAGQ 549
 DB 180 KETVISGMBLHLEIYAQRLEREYCCPCITGKPKYAFRETTIAPVPFPDTHKKQSGAGQ 239
 QY 550 YGKVIQVLEPDPEDYTKLEFSDERTFGSNIPQPFVPAVEKGLDACEGPLSGHLSGLR 609
 DB 240 YGKVIQVLEPDPEDYTKLEFSDERTFGSNIPQPFVPAVEKGLDACEGPLSGHLSGLR 299
 QY 610 FVLQDGAHMDVDSNISFIRAGEGALKQALANATLCILEPIMAVEVAPNEFOGVIAGI 669
 DB 300 FVLQDGAHMDVDSNISFIRAGEGALKQALANATLCILEPIMAVEVAPNEFOGVIAGI 359
 QY 670 NRRAGVITGDDVEDYFTLYADVPNDMFYSTTELRSCTEGGEYTMESRYOCPLESTQ 729
 DB 360 NRRAGVITGDDVEDYFTLYADVPNDMFYSTTELRSCTEGGEYTMESRYOCPLESTQ 419
 QY 730 EDVINKYLEATGQLPVKKGAKN 752
 DB 420 EDVINKYLEATGQLPVKKGAKN 442

RESULT 5
 ABB89506
 ID ABB89506 standard; Protein; 409 AA.
 AC ABB89506;
 XX
 XX 24-MAY-2002 (first entry)
 DT
 XX
 DE Human polypeptide SEQ ID NO 1882.
 XX
 XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KM antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor;
 KM vulnervary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KM cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KM neurological disease; infection; human; secreted protein.
 XX
 OS Homo sapiens.
 XX
 PN WO200190304-A2.
 XX
 PD 29-NOV-2001.
 XX
 PF 18-MAY-2001; 2001WO-US16450.
 XX
 PR 19-MAY-2000; 2000US-205515P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Birse CE, Rosen CA;

XX WPI; 2002-122018/16.
DR N-PSDB; ABL89915.
XX
XX Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT prevention of neural, immune system, muscular, reproductive,
PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT disorders -
XX
XX Class 11; SEQ ID NO 1082; 2081pp + Sequence Listing; English.
XX
XX The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC (ABL89449-ABL90444) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 409 AA:
SQ
Query Match 54.2%; Score 2113.5; DB 23; Length 409;
Best Local Similarity 98.8%; Pred. No. 3.6e-172;
Matches 405; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
QY 343 MNSSRNHPVGLAFLPVLVGRGQLTYVRSYQGLKKGITTYNTRKKVRLQRLANM 402
DB 1 MNSSRNHPVGLAFLPVLVGRGQLTYVRSYQGLKKGITTYNTRKKVRLQRLANM 60
QY 403 ADMMASTBEVYAGDICALFGIDCSGDTFTDKANSGLSMESIHVDPIVISTAMKSNK 462
DB 61 ADMMASTBEVYAGDICALFGIDCSGDTFTDKANSGLSMESIHVDPIVISTAMKSNK 119
QY 463 DLEKFSKIGRFTREDPTFKVYFDTNKEVTSIGMELHETIAORLEREYGGPCITGKP 522
DB 120 DLEKFSKIGRFTREDPTFKVYFDTNKEVTSIGMELHETIAORLEREYGGPCITGKP 179
QY 523 KVAFBETITAPVPDFTTHKKQSGAGQYGVIGVLEPLDPEEDYTKLEFSDFTFGSNIPKQ 582
DB 180 KVAFBETITAPVPDFTTHKKQSGAGQYGVIGVLEPLDPEEDYTKLEFSDFTFGSNIPKQ 239
QY 583 FVPANVEKGLDACCEKPLSGHLSGIRFLQDGAHNMVDNBSIFTRAGEALKQALANA 642
DB 240 FVPANVEKGLDACCEKPLSGHLSGIRFLQDGAHNMVDNBSIFTRAGEALKQALANA 299
QY 643 TUCIEPIIAVEVAVANEROGVIAGINRHGVTIGODGVEDFTLYADVPINDMGYST 702
DB 300 TUCIEPIIAVEVAVANEROGVIAGINRHGVTIGODGVEDFTLYADVPINDMGYST 359
QY 703 ELRSCTEGKGEYTMESRYQPCLPSTQEDVINKYLEATQQLPYKKGAKAN 752
DB 360 ELRSCTEGKGEYTMESRYQPCLPSTQEDVINKYLEATQQLPYKKGAKAN 409
RESULT 6
ID AAU35975 standard; Protein; 692 AA.
XX AAU35975;
AC XX
XX 14-FEB-2002 (first entry)
DT XX

DE Helicobacter pylori cellular proliferation protein #288.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Helicobacter pylori.
OS
XX WO200170955-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001WO-US09180.
PF
XX
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-20727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
PI Yamamoto RT, Xu HH;
PI
XX
XX WPI; 2001-611495/70.
DR N-PSDB; AAS53834.
DR
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PT
XX
XX Example 3; Seq ID No 11568; 511pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 692 AA;
SQ
Query Match 38.2%; Score 1489; DB 22; Length 692;
Best Local Similarity 45.7%; Pred. No. 2.3e-116;
Matches 319; Conservative 128; Mismatches 233; Indels 18; Gaps 11;
QY 42 PNEKIRNIGISAHIDSGTTLTERLYVYTGRLAKNHEVKGKDGAVAMDSELERGRT 101
DB 6 PLNRRIRNIGIAHIDAGTTTSSRLFTYGVSHKIGEV--HDG-AAITDMWEGEKRGIT 62
QY 102 IOSAATFTMKQDVNINIITDTEGHVDTTEVERALAVLDGAVLVLCAGVGOCOTTTVARO 161
DB 63 ITSAAITCFWMDHQINLIDTDEGHVDTTEVERSMKVLGAAVSVCVGVQPSSEITVMQ 122
QY 162 MKRYVVPFLTFINKLDKRGMSNPBALQOMRSKLNNTAFMOIPMGLBGNFKGIIVLIBER 221
DB 123 ANKYGVPRIIVFNKDKRIGANFYVNEVDIKORLKNPVPINIPJABDTFTGVIDLVQMK 182
QY 222 AIYFDGDEFSQIVRYG--EIPALRAAATDHRQELIECVANSDBQJENFLEKIPISIDL 279

```

Db      183 AIWMNNE-TWGAKYDVEIIPSDLEKAKQYREKLVBAVAEQDBALMEKYLGEELDIERI 241
Qy      280 KLAIRATILKRSSTPVLGSALEKNGVPLDAVLEYNPSPVQVYAILANKKDDSEKKT 339
Db      242 KGIKTCGLMNSVPMALCGSSFNKKGVTLLDAVIDLPAPEVNDI---KSIDRTEB 297
Qy      340 KILMNSRRNHSHPVGLAFPLEVGRF-GOLTYVRSYOGELKKDITVTRTRKVVLR 398
Db      298 EYVYKSS--DDGFALALFKIMDPVQGLTFRVYRGLKLESSYYNNTKKCKEVGL 355
Qy      399 ARHADMEASTEEVYAGDICALFGI-DCASGDTFDKANSGLSMESIHPDVISIAWK 457
Db      356 LKMSNKRRE-DIKEVAGSICAFVGLKDTLTGDLCDERKA-VLERMSEPEPEVIAIVE 413
Qy      458 PSKNDLEKRSKIGFTREDPFKYVFDTENKETVYISGMGELHLEIYQRLEREVGC 517
Db      414 PRTKAOEKKKVALGKLAEDPSFRWTOEETGLIGGELHLEIYDRLKREKVERA 473
Qy      518 ITGKPYAFRETTTAVPFDFTHKQSGAGQYKAYIGVLEPLDPEDYTLKESDETFGS 577
Db      474 EIGQPVAFRETTIRSSVSKENKAKQSGRGQYGHVFIKLEPEPG--SGYEFVNISSG 531
Qy      578 NIPKQFVPAVEKGFLLDACEKGPLSGHKLGLRFVLQDGAHHMVDNSNISFTIRAGCALQ 637
Db      532 VIPKEVYIPAVDKGIOEAMQNGVLGYPVVDFKVTLYDGSYHVDSSSEMAFKLAGMAFK 591
Qy      638 ALANATLCTLEPIMAVEVVAAPNEFOGOVYAGINRRHGVTITGQGVADYFTLVADVPLNM 697
Db      592 ASRAAPVPLEPMMKVEVEVEPEYMGVDVIGDLRRRGQINSMDRLGLKIVNAPVLEVM 651
Qy      698 FGYSLELRSCTEGKEGYTMEYSRYOPLPSTOEDVINK 735
Db      652 FGYSLELRSCATQGRGYSMSEFHDYGVSPSIATKEIYEK 689

RESULT 7
AAU35467 standard; Protein; 700 AA.
ID      AAU35467;
XX      AAU35467;
AC      14-FEB-2002 (first entry)
DT      14-FEB-2002 (first entry)
XX      Haemophilus influenzae cellular proliferation protein #108.
DB      Haemophilus influenzae cellular proliferation protein;
XX      Antisense; prokaryotic cellular proliferation protein;
XX      antidiotic; antibacterial; drug design.
XX      Haemophilus influenzae.
XX      OS
XX      PN
XX      MO200170955-A2.
XX      PD
XX      27-SEP-2001.
XX      21-MAR-2001; 2001WO-US09180.
XX      PF
XX      21-MAR-2000; 2000US-191078P.
XX      PR
XX      23-MAY-2000; 2000US-206848P.
XX      PR
XX      26-MAY-2000; 2000US-207727P.
XX      PR
XX      23-OCT-2000; 2000US-242578P.
XX      PR
XX      27-NOV-2000; 2000US-253625P.
XX      PR
XX      22-DEC-2000; 2000US-257931P.
XX      PR
XX      16-FEB-2001; 2001US-269308P.
XX      PA
XX      (ELIT-) ELITRA PHARM INC.
XX      PI
XX      Haeselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,
XX      Yamamoto RT, Xu HH;
XX      WPI; 2001-611495/70.
XX      N-PSDB; AAS53326.
XX

```

```

PT      New polynucleotides for the identification and development of
PT      antibiotics, comprise sequences of antisense nucleic acids -
XX
XX      Example 3; Seq ID No 11060; 511pp; English.
CC      The invention relates to antisense inhibitors of genes essential to
CC      prokaryotic cellular proliferation, their use in identifying the
CC      genes, their use in the discovery of novel antibiotics, the essential
CC      genes themselves and the encoded proteins. The prokaryotes used are
CC      Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC      pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC      invention is also useful for the identification of potential new targets
CC      for antibiotic development. The antisense nucleic acids can also be used
CC      to identify proteins used in proliferation, to express these proteins,
CC      and to obtain antibodies capable of binding to the expressed proteins.
CC      The proteins can be used to screen compounds in rational drug discovery
CC      programmes. The antisense nucleic acid sequence is also useful to screen
CC      for homologous nucleic acids which are required for cell proliferation in
CC      a wide variety of organisms. The present sequence represents an
CC      essential prokaryotic cellular proliferation protein.
CC      Note: The sequence data for this patent did not form part
CC      of the printed specification, but was obtained in electronic
CC      format directly from WIPO at
CC      ftp.wipo.int/pub/published_pct_sequences.
XX
XX      SQ
XX      sequence 700 AA:
XX
XX      Query Match 37.3%; Score 1454; DB 22; Length 700;
XX      Best Local Similarity 45.1%; Pred. No. 2.3e-115;
XX      Matches 318; Conservative 124; Mismatches 235; Indels 28; Gaps 13;

Qy      42 PNEIKRINIGISAHIDSGKTTLTERVLYYTGRIAKNHEVYKQGVAVNDSELEROGIT 101
Db      6 PIERFRNIGISAHIDAGKTTTERILFTYGSHKIGEV--HGC-AATDMWEGDEBERGIT 62
Qy      102 IOSAATFTWKKDVN-----INIIDTGHVDFTIEBERALRVLDGAVLVLCVAGVQCQ 154
Db      63 ITSAAITFAWSGMSQOFPQHRINVIDTGHVDFTEVERSKMLVGLGAVVYCAVGVQVQ 122
Qy      155 TMTVNRQMKRYVVPFLPTINKLDRMGSNPARLQOMRSKLTNNNTAFMOIPMGLGNPKFI 214
Db      123 SETVMRQANKYEVPRIFAVNKKDRGTANFLRVBOLKTRLGANAIPLQVPAEENFTGV 182
Qy      215 VDLIEERAI-YFDGFSQIVRYGEIPAEIRAAATRHROELIECVANSDEQLGEMLEBEKI 273
Db      183 VDLIMKALINWENAOQMTFTFEYVPAWQNDCEMRQVLVEAALASSELMEKXLGSD 242
Qy      274 PSISDLKLAIRPATLKRSFTPVFLGSALEKNGVPLDAVLEYNPSEVQVYAILANKKD 333
Db      243 LTEESEIKALRQVLANEIILVTCGSAFKNKGVQAMLDVAVVEYLPAPTDIP--AIKGINP 300
Qy      334 DSEKTKILMNSRRNS--HPVGLAFPLEVGRF-GOLTYVRSYOGELKKGDTIYNTRR 390
Db      301 DETE-----GERHASDEBPSSLAFAKIAIDPFVGNLTFPRVYSGVINSGLVLSVRQ 353
Qy      391 KKVRLQRLARHADMEASTEEVYAGDICALFGI-DCASGDTFDKANSGLSMESIHPVD 449
Db      354 KKEKRGRIQVQMANRRE-EIKEVRAGDIAAALGLDVTGDTLC-AIDAPILMEMEPFE 411
Qy      450 PVISTAMKPSKNDLEKFSKIGFTREDPFKYVFDTENKETVYISGMGELHLEIYQRL 509
Db      412 PVIASVAVPRTKADDEKKGALGRLAOEDPSFRVHTDESGETIISGMGELHLDIVRM 471
Qy      510 EREYGCPTITGKPKYAFRETTIAPV-PDFTHKQSGAGQYKAYIGVLEPLDPEDYTKL 568
Db      472 KKEFVEANIGKPYVRETTIKRTVNDVGEKAKQSGAGQYGHVVIDYLPDPDG-PGY 530
Qy      569 EESDTPGSNIPKQVPAVEKGFLLDACEKGPLSGHKLGLRFVLQDGAHHMVDNSNISFT 628
Db      531 EFWNEIKGVIPGEVYIPAVDKGIOEALGSGPLAGYPVVDIGVRLHFGVHDVDSSELANK 590
Qy      629 RAGEGALQALANATLCTLEPIMAVEVVAAPNEFOGOVYAGINRRHGVTITGQGVADYFTL 688

```


XX (ELIT-) ELITRA PHARM INC.
 PA Zyskind J, Ohlsen KL, Trawick J, Forsyth RA, Froelich JM, Carr GJ;
 PI Yamamoto RT, Xu HH;
 DR WPI; 2000-514822/46.
 DR N-PSDB; AAA65895.
 XX Novel polynucleotides and polypeptides associated with microorganism
 PT proliferation, used to identify inhibitors of bacterial growth and
 PT proliferation, for use in antisense therapy -
 XX
 XX Claim 11; Page 178-180; 316pp; English.

XX Sequence 704 AA;

XX Query Match 36.6%; Score 1427; DB 21; Length 704;

XX Best Local Similarity 44.2%; Pred. No. 4.7e-113;

XX Matches 316; Conservative 130; Mismatches 239; Indels 30; Gaps 15;

QY 42 PNEKINIGISAHIDSGKTTLTERRVYYGRIKAKMEVKGKGVGVMSMEIERGRTT 101
 DB 6 PIARYNIGISAHIDSGKTTTERRILFYGVNKHIEV--HDC-AATMMMEDEORRGIT 62
 QY 102 IGSATFTMWKOV-----NINIIDPGHVFTEVERALRYLDGAVLVLCAGVGVOO 154
 DB 63 ITSAATTAATWSGMAKQYEPHRIINIDPGHVFTEVERSMRVLDGAVLVLCAGVGVOO 122
 QY 155 TWTVNQMKRYNVPFLTFINKLDRMGSNPALQOMRSKLNHTAEMQJPMGLEGNPKGI 214
 DB 123 SETVMQANKYKVPRIAFVNMKRMGNFLKVVNQTKTRIGANPVLQALIGSEHPTGV 182
 QY 215 VDLIERRAI-YFPGDPSQIVRYGEIPEALRAAATDHRQELIECVANSDEOLGEMPLEBK 273
 DB 183 VDLVKKKAIINMDADQGVTFEYEDIPADWVELNEMHONLBSAAASEBELMEKYLGSSE 242
 QY 274 PSISDKLAIIRATLRSPFPVFLGSALKNGVQPLDLVLEKLPSESONVAIINKKO 333
 DB 243 LTAELKGAIRQVNLNNEIILVTGSAFKNGVQANLDAVIDLPSVPDP--AINGILD 300
 QY 334 DSKEKTKIIMNSSRHNS--HPFGLAFPLEVGRF--GQLTVRSYOGELKKGDTIYTRTR 390
 DB 301 DGDQTE-----ARRHNSDEPFSAALFKIATDPFVGNLTFRRYSGVANSQDVLNSVKA 355
 QY 391 KKYRLORLARMAHADMMEASTEEVYAGDICALFEI--DCASGDTFTDYANSGLSMESHIVPD 449
 DB 356 ARERFRIVOMHANKRE--EIKERVADIDAAIQLKVTTCDTLCD--PDAPILIERMEFPE 413
 QY 450 PVISIMKPSNKNDLKFSGKIGRFRDPTFKYVYEDTENKEKVISGMBELHLEIYAQPL 509
 DB 414 PVISIVPEKTKADQEKGLALRLAKLDESPFVWMBDESNQITIMGMBLHLDIIVDKR 473
 QY 510 EREYGCPCITGKPKVAFRETTITAPV--PFPETHKQSGAGQYKVGIVLEPLDPEDYTK- 567
 DB 474 KKEFNVANVGKQVAVRETIRKQYVDGSKAKQSGRGQYHVIVDWPLEPGSNPBG 533
 QY 568 LBSDBTFSQNIPIKQVPAVEKGFLLDACEKPLSGHKLSGLRFVLDDGAHMYDSNEISF 627

DB 534 YEFINDIKGVIPGEYIPAVDKIQEOLKAGLACYPVVDMGIRLHPSHYHDSSELA 593
 QY 628 IRAGEALKQALANATLCTLEPIMAVEVAPNEFOGVAGINRRHGVITGQDVEDYPT 687
 DB 594 KLAASIAFKEGKKAKPVILLEPIMKVEVETPEENTGDVIGDSRRGMKKGSESVTGK 653
 QY 688 LVADVPPLNDMFGYSTELRSCTEGKGEYTWESRYOPLPSTQEDVIYINKYLEATGQ 742
 DB 654 IHAEVPLSEMFQYATQRLSTRGRASYMEFLKYDEA--PS---NVAQAVIEARGK 704
 RESULT 10
 AAU34770
 ID AAU34770 standard; Protein; 704 AA.
 XX
 AC AAU34770;
 DT 14-FEB-2002 (first entry)
 XX
 DE E. coli cellular proliferation protein #351.
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 OS antibiotic; antibacterial; drug design.
 OS Escherichia coli.
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001; 2001WO-US09180.
 XX
 PR 21-MAR-2000; 2000US-191078P.
 PR 23-MAY-2000; 2000US-206848P.
 PR 26-MAY-2000; 2000US-207727P.
 PR 23-OCT-2000; 2000US-242578P.
 PR 27-NOV-2000; 2000US-253625P.
 PR 22-DEC-2000; 2000US-257931P.
 PR 16-FEB-2001; 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 DR WPI; 2001-611495/70.
 DR N-PSDB; AAS52629.
 XX
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX
 PS Example 3; Seq ID No 10363; 511pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes, their use in the discovery of novel antibiotics, the essential genes themselves and the encoded proteins. The prokaryotes used are Escherichia coli, Streptococcus aureus, Salmonella typhi, Klebsiella pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein.
 Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at
 ftp.wipo.int/pub/published_pct_sequences.

KM vitamin B12; bacterial infection; disease.
 OS Listeria monocytogenes.
 XX MO200177335-A2.
 PN 18-OCT-2001.
 PD 11-APR-2001, 2001WO-FR01118.
 PF 11-APR-2000, 2000FR-0004629.
 PR (INSP) INST PASTEUR.
 PA Buchrieser C, Frangoul L, Couve E, Rusniok C, Feihl H, Dehoux P;
 PI Dussauguet O, Chetoui F, Nedjari H, Glaeser P, Kunz F, Cosseart P;
 PI Daniels J, Goebel W, Krefc J, Kuhn M, Ng B, Vazquez-Boland JA,
 PI Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A,
 PI Chakraborty T, Dommann E, Hain T, Berche P, Chardit A, Durant L,
 PI Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N,
 PI Madueno E, De Pablo B, Wehland J, Kaerst U, Ertian K, Hauf J;
 PI Rose M, Voss H;
 DR MPI, 2002-010914/01.
 XX Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
 PT and prevention of Listeria and related bacterial infections, and
 PT related polypeptides -
 PS Claim 6; SEQ ID No 1108, 192pp, French.
 XX The present invention relates to the genome sequence of Listeria
 CC monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
 CC it are useful for selecting probes and primers for detecting genes in L.
 CC monocytogenes and related organisms, and for studying genetic
 CC polymorphisms and other genomes. The present sequence is a protein
 CC encoded by the genome sequence of the present invention. Proteins
 CC expressed from the genome sequence are useful for raising specific
 CC antibodies, identification of L. monocytogenes and related organisms, and
 CC for biosynthesis and biodegradation, especially biosynthesis of vitamin
 CC B12. The genome sequence and proteins encoded by it are also useful for
 CC selecting compounds that regulate gene expression and cell replication
 CC and modulate L. monocytogenes-related diseases. In addition, the genome
 CC sequence and proteins encoded by it are useful in pharmaceutical and
 CC vaccines compositions for the treatment or prevention of infections by L.
 CC monocytogenes and related organisms.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIP0
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 695 AA;
 Query Match 35.6%; Score 1387; DB 23; Length 695;
 Best Local Similarity 43.5%; Pred. No. 1.2e-109;
 Matches 302; Conservative 124; Mismatches 248; Indels 20; Gaps 10;
 Oy 44 EKIRNIGISAHIDSGKTLTTERVLYYTRIAKHEVKGXGVGVAVWDSMELEROGITTO 103
 Db 8 EKIRNIGISAHIDSGKTLTTERVLYYTRIAKHEVKGXGVGVAVWDSMELEROGITTO 64
 Oy 104 SAATFTMWDVNIINIDTPGHVDTIEVERALRVLDGAVLVLCAGVGVCOTMTVAROMK 163
 Db 65 SAATFTMWDVNIINIDTPGHVDTIEVERALRVLDGAVLVLCAGVGVCOTMTVAROMK 124
 Oy 164 RRVNPLUTINKLDRNGSNPARALQOMRSKLTNNTLPMQIPMGLBNFKIYDLIERAI 223
 Db 125 TYGVPRVNVNKKDKIGADFLVSVGTLHERLANAHPIQLPICAEDTFEGIDILIMNAL 184
 Oy 224 YFDGDSQIVRYGEIPAELEAAATDHRQELIECVASDIDGLGMPLEEKPSISDKLAI 283
 Db 185 YIEDDIGNBPHIETPADLKDLDYRGLKVENVAELDELENNKYLEGEEITTEELIAGI 244
 Oy 284 RRATLRSTPVPFLGSLAKNKGVQPLDAVLLEYLPNPSVQVYAIINKKDSKEKTIIM 343

Db 245 RKGTIANVEFYPVVCGTAFKNGVQPMILDVAVLDYLPAPTDVP--AINGVLPDGE----- 296
 Oy 344 NSRRH--NSHPVGLAFLEVGRF--GOLTVVSYGELKKGDTIYTRTRKKVRLQRLAR 400
 Db 297 -AARHADSEPPSSLAFFKMTDPYVGRLTFFPVSGTLNNGSSVQVSTGKKEVGRITIQ 355
 Oy 401 MHADMMEAETEEVYAGDICALFGI--DCASGDTFTDKANSLSMESIHVPDVIAMKPS 459
 Db 356 MHANRE--EISIVYAGDIAAAGLQDTTGTGLDCEKEQ--ILLESNFFEPPIYQVAILRK 413
 Oy 460 NNDLEKFSKGIQGFREDPTFKVYFDTEKNTVIGSGELHLEIYAQLEBRVYCCPT 519
 Db 414 SKADQKMGQALAKLALEDPTFRAEFDQGTGLISGMELHLDILYDHRREPFVEANY 473
 Oy 520 GKPKVAFRETTIAPVPFDTHKKSGAGAGYKVGIVLEPLBEDYTKLEFDETFGSI 579
 Db 474 GPOVSYRETFPKSAQVSGKRVQSGRGQYGHVWIEGPP--NEBKGFEFENAVIGSV 531
 Oy 580 PKQFVPAVEKGLDACEKPLSGHKLGLRFLYLDGAAHMMVDSNETISFTRAGEALKOL 639
 Db 532 PREYIPAVQAGLEGALDNGVLAGYPLIDIKAKLYDSHYDVSNEVAFVAAASMLARNA 591
 Oy 640 ANATLCIEPIMAVEVADNEFQGVYIAGINRRHGVITGQDVEDYFTLYADVPLNDFG 699
 Db 592 KKCDPVLIEPMAAVEVIEEYVIGDIMGNITSRRGRVDGMBARGNAQVRAFPVPLANMF 651
 Oy 700 YETELRSCREKGEYTMEXSRQPCLPSTQEDVI 733
 Db 652 YATHLRSGTQGRGVYTMQFDHYEVPKSIABEII 685
 RESULT 14
 AAU38372
 ID AAU38372 standard; Protein, 704 AA.
 AC AAU38372;
 XX
 DT 14-FEB-2002 (first entry)
 XX
 DE Salmonella typhi cellular proliferation protein #263.
 XX
 KW Antisense; prokaryotic cellular proliferation protein;
 KM antibiotic; antibacterial; drug design.
 XX
 OS Salmonella typhi.
 XX
 PN WO200170955-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 21-MAR-2001, 2001WO-US09180.
 XX
 PR 21-MAR-2000, 2000US-191078P.
 PR 23-MAY-2000, 2000US-206848P.
 PR 26-MAY-2000, 2000US-207727P.
 PR 23-OCT-2000, 2000US-242578P.
 PR 27-NOV-2000, 2000US-253625P.
 PR 22-DEC-2000, 2000US-257931P.
 PR 16-FEB-2001, 2001US-269308P.
 XX
 PA (ELIT-) ELITRA PHARM INC.
 XX
 PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 PI Yamamoto RT, Xu HH;
 DR N-PSDB; AAS56231.
 DR MPI, 2001-611495/70.
 PT New polynucleotides for the identification and development of
 PT antibiotics, comprise sequences of antisense nucleic acids -
 XX Example 3; Seq ID No 13965; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Bacteriella coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 704 AA:

Query Match 35.2%; Score 1374; DB 22; Length 704;
Best Local Similarity 42.6%; Pred. No. 1.6e-108;
Matches 306; Conservative 126; Mismatches 250; Indels 36; Gaps 15;

42 PNEKINIGISAHIDSGKTLTERVLYTG---RIAKMEVKKDGGAVMDSMELEKOR 98
6 PIRKRNIGISAHIDSGKTLTERVLYTG---DGAATMDMEQDER 59
99 GTTIGSAFTWMDV-----NINIDTGHVDFTEVERALRVLDGAVLCAVGV 151
60 GITXTSATATAPFSGMAKQVEPRHINIDTGHVDFTEVERSMRVLDGAVVYCAVGV 119
152 QCGTMTVNRQMKYNNPFLTFINKLDRMGSNPARAQONRSKLNHTAFMOIMGEGNF 211
120 QPQSEVTVNRQMKYNNPFLTFINKLDRMGSNPARAQONRSKLNHTAFMOIMGEGNF 179
212 KGVLDLIEERAI-YPDGDSQIVRGEIPALRLAANTDRHOELIECVANSDELGMEL 270
180 TGVVDLVKMAIMWMDQGVTEYEDIPADWDLNEMHQNLIESAAESSELMEXYLD 239
271 EKIPISIDSLKLAIRATLKRSTFPVFLGSAKKNQVOPLDVALEYLPNSEVQNTAAILN 330
240 GELTSEEBEIKQALRQVRLNEILLVCGSAFKKQGMADNVIDILPSVDVP--AING 297
331 KQDSKEKTKILNNSRHNS--HPVGLAPLEVGRF-GOLTYVRSYQEGKKGDTTYVT 387
298 ILDDGKDTP-----AERHASDDEPFSAIAFKIATDPFVGVLTFPRVYSGVNSGDTVLNS 352
388 RTRKAYRLORLARMHADMEASTEEVYADICLFGI-DCASGDTFIDKANSGLSMESIH 446
353 VKTAREBFHIVQHNANKRE-EIKERAGDIAAIGKDVTTGDTLCLD-PEXITIEKHE 410
447 VPDVVISIANKPSNKDLEKFSKIGRFTREDPTFKYVPFTENKETVVISGMEHLIEYA 506
411 LPEPVISXAVEPKTKDXXKGLALGRLAKEDSPFRWDEBNQITIAEMGLHIDILY 470
507 QRLREYGCPCITGKPVAFRETTITAV-PFDTHKKQSGAQYGVIVLEPLDEPY 565
471 DRMKREFFNRYANQKQVAYRAIRAKVTDIEGKHAQSGRGQYGHVVIDMYPLRPGSN 530
566 TK-LFSRDEFFGNSINPQFPAVEKGFGLDCEKPSGHLISGLRFLVLDQGAHMDNSN 624
531 PKQYETINDKGVIGEEYIPAVDKIGQBLKSGPLAGYVVDLGVRLHFGSYHDDSS 590
625 ISFIRAGEGALKOALNATLCILEPIMAVEVAPNEFOGVINGINRHGVIGQDVED 684
591 LAKFLAASIAFKSGFKAKAVLEPIMKVEVTEPENTGVIGDLSRRKMLMGQSESVT 650
685 YFTLVADVPLNDMFGSTELRSCTEGKGEYTMESRYQCLPSTQEDVINKYLEATGQ 742

DB 651 GVXIHAEVXLXEMFGYAXQXRSXTXGRASVTMKFLKYD---XAPXYVAQAVIEARGK 704

RESULT 15

ID ABP30759 standard; Protein, 692 AA.

AC ABP30759;

DT 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 10694.

KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

PN WO200234771-A2.

PD 02-MAY-2002.

PF 29-OCT-2001; 2001WO-GB04789.

PR 27-OCT-2000; 2000GB-0026333.

PR 24-NOV-2000; 2000GB-0028727.

PR 07-MAR-2001; 2001GB-0005640.

PA (CHIR-) CHIRON SPA.

PA (GENO-) INST GENOMIC RES.

PI Telford J, Masigant V, Margart R, Grandi G, Fraser C;

PI Telford H;

DR WPI, 2002-352536/38.

DR N-PSDB; ABN71390.

XX New Streptococcus protein for the treatment or prevention of infection

PT or disease caused by Streptococcus bacteria, such as meningitis, and

PT for detecting a compound that binds to the protein -

XX Claim 1; Page 4179; 4525BP; English.

XX The invention relates to a protein (ABP25413-ABP30895) from group B

CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS

CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in

CC the specification. The proteins have antibacterial and antiinflammatory

CC activity. (I), nucleic acid encoding (II), ABN6044-ABN71526 and

CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.

CC Nucleic acid encoding (I) are used to detect Streptococcus in a

CC biological sample. (I) is used to determine whether a compound binds to

CC (I). A composition comprising (I) or a nucleic acid encoding (II), may be

CC used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic

CC acid encoding (II) may be used to recombinantly produce (I) and may be

CC used in gene therapy. Antibodies to (I) are used for affinity

CC chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

SQ Sequence 692 AA:

Query Match 35.2%; Score 1371.5; DB 23; Length 692;
Best Local Similarity 42.7%; Pred. No. 2.6e-108;
Matches 296; Conservative 131; Mismatches 252; Indels 15; Gaps 8;

44 EKIRNIGSAHIDSGKTLTERVLYTGRIAGHEKQDGVAVMDSMELEKORIGITIQ 103
8 EKIRNIGSAHIDSGKTLTERVLYTGRIAGHEKQDGVAVMDSMELEKORIGITIQ 64
104 SATFTWMDVNNINIDTGHVDFTEVERALRVLDGAVLCAVGVQCGTMTVNRQMK 163

```

Db      65  SAATTAQWMDGHRVNIIDTGHVDFTEVORSLFVLDAVTVLDAQSGVEPQETETVWRQAT 124
QY      164  RYVNPFLTFTINKLDRMGSPARALQOMRSKLNNTAEMOI PMGLBGNPKGIYDLIEERAI 223
Db      125  EYGVPRIVPANKNDKIGADFLYSVQSLHRLQANNAHPIO LPISEDDFRGIIDLIKMAE 184
QY      224  YPDGDFSQIVRYGELPAELRAAATDRQELIECVANSDEQLGEMFLEEKI PSISDLKLA 283
Db      185  IYTNLDGTDILBEDIPAEYVDQANEYREKULVEAVADTDEDLMKYLEGEBETBELMAAI 244
QY      284  RRATLKRSFTVPVFGSLANKKGVQPLDAVLEYLEPNPSEVQNTAII LKKODSKEKTKIIM 343
Db      245  RKATINVEFPVLCGSAFKKGVQLMLDAVIDYLPSPLDI PAIKGINPDTDEE----- 298
QY      344  NSGRHNSHPFVGLAFLVGRF-GOLTYVRSYOGELKGDITYNTR TRKKVRLQRLAHM 402
Db      299  TRPASDEBPALAFKIMTDPFVGRLPFRKYSGLVNSGSYV LNTSKGKERIGRILOMH 358
QY      403  ADMMEASTEVEVYAGDICALFGI-DCASGDTFTDKANSGLMESI HVPDPVISIAMPKSNK 461
Db      359  ANSRQ-EIETVYAGDIAAAGVGLKDTTGSLSLTDK-KSKV LBSIEVBPVIOLMVEPKSK 416
QY      462  NDLEKFSKGI GRFTREDPTFKYFTDENKETVISGMBELHEIY AORLEBERGCPCTGK 521
Db      417  AQQDKKGIALQKLABEDPTFRVETNETGETVISGMGELHLDV LDRMKKEFFVEANVGA 476
QY      522  PKVAFRETTAPVPFPDFTHKKSGAGOVKVICLEPLDEPTYK LEFSDTEFGSNIRK 581
Db      477  POUVSRETTPRASTOARGFFRRQSGKQFGDV--WLEFTN EBGKGFEPENAIWGVVPR 534
QY      582  QFVPAVEKGLDACEKGPLSGHLSGLRFLVLODAHMYDSNEI SFIRAGEGALKOALAN 641
Db      535  BPIPAVEKGLVESMANGVLAGYPMVDYKAKLYDGSYHDVDS SETAFKIASLALKEPAKS 594
QY      642  ATLCLIEPIMAVEVAPNEFQGOVIAGINRRHGVITGQDGVEDY FTLVYADVPIINDMFGYS 701
Db      595  ACPAILEPMLVTITAPEDNLAGDMGHTARGRVDMGEMAR GNTQVVRAFPVLAEMFGYA 654
QY      702  TELRSCTEGEGEYTMESRYQPCLPSTQEDVINK 735
Db      655  TVLRSATQGRGTFPMVFDHYEDVPKSVQEBIIRK 688

```

Search completed: July 14, 2003, 18:15:02
 Job time : 30.0723 sec

THIS PAGE BLANK (USPTO)

QY 396 ORLARNHADMMEEASTEEVYAGDICALFGI-DCASGDTFTDKANSGLSMESIHPDPVISI 454
DB 365 GRLLQHNANSRQ-EIOTVYSGELIAAVALGKETGTOTLCEKND-IILSMEEPEEVIHL 422
QY 455 AMKPSNKNDLEKFSKIGRFTREDPTFKVYFTDENKETVYISGGEHLLEIYAORLEREY 514
DB 423 SVEPKSKADODKMTQALVKLQEDPTFHATDETEQVILIGGGEHLILILVDRMKERN 482
QY 515 CPCTKPKVAFREBITTAPVPFDTFKHKSGGAGQYKVLGVLEPIDPEDYTLERSDET 574
DB 483 VECNVGAPMVSYETKRPQAPQVQKSRSGRGQYGDV--HLETPNETGGGFEBENAI 540
QY 575 FGSNPKOPVPAVEKGFPLDACCKEPLSGHLSGLRVLQDGAHMYDSNEISFIRAGEGA 634
DB 541 VGGVREVIYPSVEQGLKAMENGVALGYPLIDVKKALFDGSHYDVDSSEMAFKIAASLA 600
QY 635 LKALANATLCTLEPTMAVEVAPNEQGOVINGIRRHGVITGQGVEDYFTLVADVPL 694
DB 601 LKEAARKCDPVILEPMMKVTIEMPEBYMGDIMGDVARRGRVDMGEMPRGNAQVYNAVPL 660
QY 695 NDMFGYSTELRSTCEKGEYTMESRYQPCLPSTOEDVINK 735
DB 661 SEMFGYATSLRSNTQGRGYTMTFPHYAEVPSIAEBIILK 701

RESULT 2

US-07-718-535-3
Sequence 3, Application US/07718535
Patent No. 5322784
GENERAL INFORMATION:
APPLICANT: Salyers, Abigail A.,
APPLICANT: Shoemaker, Nadja B.,
APPLICANT: Nikolich, Mikeljon P.,
TITLE OF INVENTION: Method and Materials For
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: William Brinks Olds Hofer Gilson and Lione
STREET: P.O. Box 10395
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60610
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM XT
OPERATING SYSTEM: MS-DOS 3.31
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07718,535
FILING DATE: 05-JUN-1991
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Crook, Marnell W.,
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3617/22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 321-4200
TELEFAX: (312) 321-4299
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 641 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown

Query Match 14.6%; Score 569.5; DB 1; Length 641;

Best local similarity 26.1%; Pred. No. 212e-47;

Matches 182; Conservative 120; Mismatches 287; Indels 109; Gaps 17;

QY 46 IRNIGISAHDSKTLTTERVLVYTGRIAMGEVKKDGVGAVMDSMELEROGITIGA 105

DB 3 IINLGLAHIDAKTSVTENLLFASGATEKCGCVNDG---TITDSMIEKRRGITVRAS 59
QY 106 AFTFMKOVNINIIIDTPGHVDFTIEVERALRYLDGAVYLCAVGVGQOTMVRNOMKRY 165
DB 60 TTSIIMNGKCAIIDIIPGHMDPIAEVERFKPLDQAVILLSKESIQAOIKLFTTLQTL 119
QY 166 NVPEPTINKLDRMGSPAPALQOMRSKLNHTAFMOJPMGLEGNFKGIVDLIBERAIYF 225
DB 120 QLPITLFIINKIDRAGVNLERLTLDKANLSQVLFMO-----NVVD-----GSYV- 164
QY 226 DQDFEQIYRGERIPALPAAATDRHOELIECVANDEQLEMEFLEKIPSLISDLKLAIR 285
DB 165 -----PV-----CSQTYIKEEYKEFVCNHDNDILERYLADSEISPDYNNYITIA 208
QY 286 ATLKSPFPVPLGSLAKKKGVOPPLDAVLEYLPNPSVONVAILINKKDSKEKTKILMNS 345
DB 209 LVAKAKVYFVLHGSAMFNIGINELLDALITFLLPAPASNRL-----SLLYLIENHP 261
QY 346 SRHNSHPVGLAFPLEVGRFGQLTYVRSYOGELKKGDTIYNTTRKKVRLQRLAMHADW 405
DB 262 KGH-----KRSFLKIIDGSLRLDVDVRINDSEKFIKIKVLTINQ- 302
QY 406 MEASTEEVYAGDICALFGID-----CASGDTFTDKANSGLSMESIHPDPVIS 453
DB 303 REINDEVGANDIAIVEDMDPRIGNYLAEPCLL-----OGLSHQ-----HPLK 348
QY 454 IAMKPSNKNDLEKFSKIGRFTREDPTFKVYFTDENKETVYISGGEHLLEIYAORLEREY 513
DB 349 SSVRPDRPERKSKVYLSALNTLMIEDPSLSFINSISDELSLYLTQKEIITLLEERF 408
QY 514 GCPCTITGKPKVAFRE-----TITAPVPFDTFKHKSGGAGQYKVLGV-LEPLDEP 564
DB 409 SVKHFDEIKTIYKRPVKKVKIKIIEVP-----PNFYWATIGLTLEPLIGT 457
QY 565 YTKLERSDPTFGSNIPKQFVPAVEKGFPLDACCKEPLSGHLSGLRVLQDGAHMYDSNE 624
DB 458 GLQIR-SDISYG-YLNHSFONAVPEGIRMSCSG-LHSMEEVTDLKVTFTOAEVSPVSYP 514
QY 625 ISFIRAGEGALKQALANATLCTLEPTMAVEVAPNEQGOVAGINRRHGVITGQGVED 684
DB 515 ADFRQUTPVYFRFLAQSGVDLLEPMLYFELQIPQASSKAITDQKMSLEIDISCNNE 574
QY 685 YFTLVADVPLNDMFGYSTELRSTCEKGEYTMESRYQ 722
DB 575 WCHIKGVPLNTSKOYASVSYTGTGLGHPMKPCGYO 612

RESULT 3

US-08-161-999-3
Sequence 3, Application US/08161999
Patent No. 5674733
GENERAL INFORMATION:
APPLICANT: Salyers, Abigail A.,
APPLICANT: Shoemaker, Nadja B.,
APPLICANT: Nikolich, Mikeljon P.,
TITLE OF INVENTION: Method and Materials For
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSER: William Brinks Olds Hofer Gilson and Lione
STREET: P.O. Box 10395
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60610
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
COMPUTER: IBM XT
OPERATING SYSTEM: MS-DOS 3.31
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/161,999
FILING DATE: 02-DEC-1993

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/718,535
FILING DATE: 05-JUN-1991
ATTORNEY/AGENT INFORMATION:
NAME: Crook, Marnell M.
REGISTRATION NUMBER: 31,071
REFERENCE/DOCKET NUMBER: 3617/22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 321-4200
TELEFAX: (312) 321-4299
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 641 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
US-08-161-999-3

Query Match 14.6%; Score 569.5; DB 1; Length 641;
Best Local Similarity 26.1%; Pred. No. 2,2e-47;
Matches 182; Conservative 120; Mismatches 287; Indels 109; Gaps 17;

QY 46 INNIGISAHIDSGKTLTERVLYYGRIAKMEVKGKGVGAVMDSMEIEROGITIGSA 105
DB 3 INNLGIAHIDAGKTSVTENLLFASGATEKCGCVNDG---TTDSMDIEKRGITVTRAS 59
QY 106 APTFMKQDVNINIDTGRHVDTIEVERALRVLDGAVVLCANGVOCOTMTVNRQMKRY 165
DB 60 TSIITWNGVKCNITDPGHMDTAEVETPKMLDGVALLISABEGIOAQOTKLLFNTLQL 119
QY 166 NVPLTFPINKLDRMGSNPAPALQOMRSKLNHTAFMQIPMGLEGNFKGIIVDLIEEPAIF 225
DB 120 QPTTIFINKIRAGNLERLYDIKANISQDVLFWQ-----NVYD---GSYV- 164
QY 226 DGFSGQIYVGEIPAEPLRAATDROELIECVANSDEQLSEMFLEKIPISIDLKAIIR 285
DB 165 -----PV---CSQTYIKSEYEFVCHNDNIEERYLADEISPADYNTIILA 208
QY 286 ATLKSGFTFPLGSAKKNKGVOPLLDAVLEYLPNSEVQNVAILNKKDSKEKTKILMNS 345
DB 209 LYAKAKVYVVLHGSAMFNIGINELDAITSFILPPASVSRL-----SSVYKLEHDP 261
QY 346 SHNSHPVGLAFPLEVGRFGQLTYYRSYQGELEKKGDTIYNTTRKKVRLQRLARHADM 405
DB 262 KGH-----KRSFLKIIDGLRLADVAINSEKFTIKULKTINOG- 302
QY 406 MEASTEEVYAGDICALFGID-----CASGDTFTDKANSGLSMESIHPDPVIS 453
DB 303 REINDEVGANDIAIVEDMDDFRIGNVLAEPCLL-----QGLSHQ-----HPALK 348
QY 454 IAMPKRNKNDLEKFSGIGRFTREDPTFKYPTTEKKEVVISGMGELHLEIYARLEREY 513
DB 349 SSVRRPRPERSKVISALNTLWIEDSLFSGINSYDELEISLYGTOKEIITQTLLEBRP 408
QY 514 GCPCTIGKPKVAFRE-----TITAPVPDFTHKSGAGAGQYKVIQV-LEPLDPD 564
DB 409 SVKVNDELKITYKEKPKVKKKIKIQLIEVP-----PNPYATIGLTIEPLPLGT 457
QY 565 YTKLEFSDTEFGSNIPKQFVPAVEKGFLLDACEKPLSGHKLGLRFLVDGAHHMVDNS 624
DB 458 GLQIE-SLISYSG-YLNHSFQNAVFEGIRMSQSG-LHGWEVTDLKVTFQAEYISVSYF 514
QY 625 ISPIRAGEALKQALANATLCTLEPITMAVEVAPNPFQCVYAGINRRRGVITGQGVGD 684
DB 515 ADPRQLTTPYVFRALAQOQGVLDLEPMLYELQIPQAASEKAITDLOKMSIEDISCSNNE 574
QY 685 YFTLYADVPLNDMFGYSTELRSCTEGKEGYTMEYSRYQ 722
DB 575 WCHIKGVPLNTSKOYASBVSYYTKGLGIFMWKPCGYQ 612

RESULT 4

US-09-293-549-2
Sequence 2, Application US/09293549
Patent No. 6440409
GENERAL INFORMATION:
APPLICANT: G. Todd Milne
APPLICANT: Gerald Fink
TITLE OF INVENTION: A METHOD FOR EFFICIENT AND HIGHLY
TITLE OF INVENTION: SELECTIVE CONTROL OF MICROORGANISMS
FILE REFERENCE: 50078/008002
CURRENT APPLICATION NUMBER: US/09/293,549
CURRENT FILING DATE: 1999-04-16
EARLIER APPLICATION NUMBER: 60/082,089
NUMBER OF SEQ ID NOS: 14
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 842
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-293-549-2

Query Match 13.9%; Score 541.5; DB 4; Length 842;
Best Local Similarity 23.7%; Pred. No. 2,2e-44;
Matches 201; Conservative 133; Mismatches 278; Indels 235; Gaps 30;

QY 46 INNIGISAHIDSGKTLTERVLYYGRIAKMEVKGKGVGAVMDSMEIEROGITIGSA 105
DB 19 VANNSEVIAHVDHSGKTLTDSLVQAGIIS-----AAKGAEAFPTTRKDEGRGIIIST 73
QY 106 APTFMKQDVN-----INIDTGRHVDTIEVERALRVLDGAVVLCVAV 148
DB 74 A-ISLYSEMSDEVDKIKQKTDGNSFLINLIDSFGHVFSSEVTAALRVTDGALVVDPIT 132
QY 149 GGVOCOTMTVNRQMKRYVNPFLTFINKLDRMGSNPAPALQOMRSKLNHTAFMQIPMGLE 208
DB 133 EGVCVQTEVTLKQALGERIKPVPVINKVDR-----ALLEKQVSKE 172
QY 209 GNFKGIIVDLIEERAI-----YFPGDPSQIYVGEIPAE-----LRAAATDH 249
DB 173 DLYQFARTVESVNVIVSTYADEVDQVY---PARGVAFSGSLHGMAFTRIPARTY 229
QY 250 ROEL-IECVANSDEQLSEMFLEKIPISIDLKAIIRATLKSGFT-----PVF-LGSAIK 302
DB 220 AKKFGVDKAKMMDRLMGDSFFNPRTKWTNKOTDAEGKPLERAFNNFILDPFRLFTALM 289
QY 303 N-----KGVQ-----PLDAYLE---YLPPPSRYQ 324
DB 290 NPKKDEIPVLEKLEIVLKGDEKLEGKALLKVNRKFLPADALLEMIVLHLPSPVTAQ 349
QY 325 NY-----ATLNKKDSKEKTKILMNSRRNSHPFVGLAFPLEVGRFGQLTYY 370
DB 350 AYRAQOLYBGPADNANCAIKACDPKADLMLYSKVFTS-----DKGRF-YAF 397
QY 371 VASVQGELEKKGDTIYNTTRKKVR-----LORLARHADMMEASTEE 412
DB 398 GAVFAGTYKSG-----QKVRIGPNVYVPGKKDPLFIKAIQGVLMGMGFVE-PIID 447
QY 413 VYAGDICALFGID---CASGDTFTDKANSGLSMESIHPDPVISIAMPKRNKNDLEKRSK 469
DB 448 CPAGNIIGLVGIDQPLTLTGTLTSETAHNNKVMKFSV-SPVQVAVBRKANNDLPKQVE 506
QY 470 GIGRFTREDPTFKYPTTEKKEVVISGMGELHLEIYARLEREY-GCPCTIGKPKVAFRE 528
DB 507 GIKRISKSDPCVLTWM-SBSGHEIYAGGELHLEICLDLLEHDAHCVPLKISPPVAYNE 565
QY 529 TITAPVPDFTHKSGAGAGQYKVIQVLEPLDPDYTKLSE----- 569
DB 566 IYES-----ESSQATALSKSPNKNRNIYLAABPIDEVSLAIENGINKPRDDFARARIMAD 621
QY 570 -----FSDTEFGSNIPKQFVPAVE-----KFLDACEKPLSGHKL 605
DB 622 DYGMDVTDARKIWCFGPDGNGPNLVIDQTKAVQYVLAHRIKDSVVAFAWATKGPITGSEEM 681

Matches 150; Conservative 78; Mismatches 181; Indels 97; Gaps 15;

```

QY 44 EKIRNIGISAHDGKTTLTERRVLYTGRKAKHEVKG-KDGGAVWDSMELEKRGIT 102
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 ESKRTAIIISHPAGKTTITTEKLLTSGAIREAGYKCKTGKTFASIDMKVAQEGISV 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 103 QSAATFTMKDVNINIIDTPGHVDFTEVERALRVLDGAVLYCAVGVOCOTMTVNRQM 162
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 62 TSSVMQPDVDYKINILDFPGHEDFSDTYRTLMVDSAMVLDCAKGIETPKLKFVC 121
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 163 KRVNPELFTINKLDMGNSPARALQOMSKLNHTAFMOIPWLGKNGFGIYD----- 216
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 122 KMGKIFFTFINKLDMVGKPEFELDEIETLIDTYPMWPMVGOMQNFEGIIIDRSKTI 181
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 217 -----LIBERAIYFDGDFSOIVRGEIPAELEAATDHRDELICVAN 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 182 EPPRDEENLHNEDELKEEHAIKNDSAFEOAI-----EEMLV--- 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 SDEQLGEMPLEEKIPISIDKLAIIRATLKRSFTPVFLSGALKKGVOPLLDAVEYLPN 319
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 222 --DEAGAFDNE-----ALLNGELTPVFGSALANFGVQNFANAYVDHAM 265
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 320 PSVQVYALINKKDSKEKTKIIMSSRHNSHP---FVGLAPLEVQ---RFGQLTYV 371
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 266 PNARQ-----TKEEVVDV-----SPFDIDFGSGFTFKIQANMDPKHRIRIAM 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 372 RSVQSELKGGDTIYNTRTKRVRLORLARHADMEASTEEVYAGDICALFGI--DCASGD 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 307 RVVSGAFERMDVTLQRTNKQKQITSTSPNADKETVNHAV-AGDILGYDTGNQIQIG 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 TPTDKANSGLSMESI-HVPDPVISIAMKPSKNKDLK--FSKGIQGFREDPTPKYVPT 487
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 366 TLV---GGKQKVSFQELPQFTEIFMKVSAKVMQKQHFHKGIEQLVQEG-AIQYKTL 420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 488 ENKETIYISGGEHLHETIYQRLEREY 513
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 421 HTNQIILGAVGQLQFVEFHEHNMNEY 446
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 9

```

US-08-984-618-9
; Sequence 9, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TBLX: 133521

```

INFORMATION FOR SEQ ID NO: 9:

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 527 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Haemophylus influenzae

```

US-08-984-618-9

Query Match 11.7%; Score 456.5; DB 4; Length 527;

Best Local Similarity 28.7%; Pred. No. 2.6e-36;

Matches 144; Conservative 82; Mismatches 211; Indels 65; Gaps 14;

```

QY 45 KIRNIGISAHDGKTTLTERRVLYTGRKAKHEVKGKDVG-AAMDSELEKRGITQ 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 10 KRTTAAIISHPAGKTTITTEKLLYGNALQTGSAVKGSAHAASDMMEMEKORGISIT 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 SAATFTMKDVNINIIDTPGHVDFTEVERALRVLDGAVLYCAVGVOCOTMTVNRQM 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 70 TSSVMQPDVDYKINILDFPGHEDFSDTYRTITAVDSCLMVIDSAKVEERTIKLMEYTR 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 164 RYNVPELFTINKLDMGNSPARALQOMSKLNHTAFMOIPWLGKNGFGIYDLI-EERA 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 130 LADTPIIFPMKNDLRDINDPIELDEVENVLKIRCAPITWPIGCKKLFQGVYHLAKDETY 189
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 223 IYFDGDFSOI---VRGEIPAELEAATD-----HROELIECVANSDEQLGEMFLEE 271
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 190 LYQSGSGSTIQAVRVKGLNDELVAAGDDLAQQLREBELVQASNEFEQ----- 241
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 272 KIPSIDKLAIIRATLKRSFTPVFLSGALKKGVOPLLDAVEYLPNSEVQNYAIIINK 331
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 242 -----DAFIKELTPVFGTRLGNFGVDHFDLGTQNAFKQSQHQ--ADRT 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 332 KDSSEKTKIIMSSRHNSHPVGLAPLEVQ---RFGQLTYVRSYQELKGGDTIYNT 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 287 VESAEEK-----FSGFVFKIQANMDPKHRVAFMIVSGYKGMKLKIV 332
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 RTRKRVRL-QRLARHADMEASTEEVYAGDICALFGIDCAS-GDTPTDKANSGLSMESI 445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 333 RIGKDVISDALTFMAGD--RAHAEVAYAGDIIIGHNHTQIGDTFT---QGETLKFT 386
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 446 HVPDPVISIAMKPSKNKDL--EKFSGIGIRFREDPTPKYVDTNKETVISGEMELHLE 503
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 387 GIPIAPBELFRIRLKDPLKQKQLKGLVQLS-EEGAVQVFRPLNNDIIVAGVGLORD 445
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 504 IYQRLEREYGCPCITGKPKVA 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 446 VVVSRLKTEYNVEALYENNVVA 467
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 10

```

US-08-984-618-7
; Sequence 7, Application US/08984618
; Patent No. 6251647
; GENERAL INFORMATION:
; APPLICANT: de Lencastre, Herminia
; TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
; TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue, 4th Floor
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/984,618
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-213 CIP
; TELEPHONE: 201-487-5800
; TELEFAX: 201-343-1684
; TBLX: 133521

```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,618
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-213 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 517 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: *Staphylococcus aureus*
US-08-984-618-7

Query Match 11.4%; Score 445.5; DB 4; Length 517;
Best Local Similarity 28.0%; Pred. No. 3.1e-35;
Matches 143; Conservative 78; Mismatches 179; Indels 111; Gaps 19;

44 EKIRNIGISAHIDSGKTLTERVLYYTGRIAKHVEYKGDVG-AVWDSMELRQGITI 102
8 EARKTPAIIISHPDAGKTTITKLVYSGAIRKAGTVKGLVNLRYVMKTKVEQEGISV 67
103 QSAATFTMKDVNINIIDTPGHVDFTEVERALRVLDGAVLV-CAVGVQOCOTMTVNRQ 161
68 TSSVMQFDVDDYELNLDTPGHEDF--EDYRTLMAVDSAVMYIDCAKQ---VEPLTFKY 122
162 MKRVNVPFLTFINKLDRMGSNPBALQOMRSKLNHTAFMQIPMGEGNFKGIVD----- 216
123 CMKRGITPITFINKLDRVGEKEPELIDETETLNTETTYMNPFGOSFGIIDRKSKT 182
217 -----LIBERAIYFDGDFGQIVRYGRIPAELRAATDRHOELIECYA 258
183 IEPFDEENILHNDPFEEDHATINDSDFQAI-----BELMLV-- 223
259 NSDEQGEWFLEBKIPISIDKLAIIRATLKRSFTPVFLGSAKKNKGVOPLLDAVLEY-- 316
224 ---EEGGEAFDND-----ALLSGDLTFVPFGSALANFGVQNFNAAYVDFAF 266
317 LPNPEVQVYALINKKDSKEKTKILMSSSRNHSHPVGLAFPLEVG---RFGQLTYVR 372
267 MKNARQTKNNVNSPFDSDS-----FSGFIRKIQANMDPKHRDRIAFMR 309
373 SYQGEIKK-GDTIYNTRTKRVLRQRLARMAH-----DWMKEASTEVEVADICALFGI-D 425
310 VVSGAFERVMMLCNVLKSK-----RSHVQRIHQIITIKLVNNAVAGDIIIGLYDTGN 362
426 CASGDTFTIKANSGLSWESIH-VPRPVISIAMKPSKNKDLK--FSKGIQRTREDPTFK 482
363 VOIGITLV---GKQATYSFODLPQTPREIFMKVSAKNVKKQKHFKGIEQLVQEG-AIQ 417
483 VYFDEENKTVISGGEHLLEIYAORLEREY 513
418 YKKTHTNQIILGAVGQLQFVFEHMKQVEY 448
Db

RESULT 11
US-08-984-618-10
Sequence 10, Application US/08984618
Patent No. 6251647
GENERAL INFORMATION:
APPLICANT: de Lencastre, Herminia

APPLICANT: Tomasz, Alexander
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,618
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-213 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 529 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: *E. coli*
US-08-984-618-10

Query Match 11.3%; Score 441; DB 4; Length 529;
Best Local Similarity 28.2%; Pred. No. 9.1e-35;
Matches 147; Conservative 89; Mismatches 222; Indels 64; Gaps 15;

45 KIRNIGISAHIDSGKTLTERVLYYTGRIAKHVEYK-DEVGAMDSMELRQGITIQ 103
12 KRRTPAIIISHPDAGKTTITKLVYSGAIRKAGTVKGLVNLRYVMKTKVEQEGISV 71
104 SAATFTMKDVNINIIDTPGHVDFTEVERALRVLDGAVLV-CAVGVQOCOTMTVNRQWK 163
72 TSSVMQFPYHDCLVNLDTPGHEDFSEDYRRTITAVDCCLMYIDAKGVEDRTKLMETVR 131
164 RYNVPFLTFINKLDRMGSNPBALQOMRSKLNHTAFMQIPMGEGNFKGIVDLIBRAI 223
132 LRDPTILTFMKNKLDRIDIDPMEILDEVENEKIGCAPITWPIGCGKLPKGVNHLKXDety 191
224 YFD---GDFQIVR--GEIPAELEAA-----ATDRQELIECVANSDEQGEWFLEBK 273
192 LYQSGKHGHTIOWRIVKLNPNDDIAVAGEDLAQQLRDELELVKASNPEDKELLAGET 251
274 PSISDKLAIRATLKRSFTPVFLGSAKKNKGVOPLLDAVLEYLPNPEVQVYALINKKD 333
252 -----TFVFGTALNGVDMGLGVEMARF-----MPRQT 284
334 DSKERTKILMSSSRNHSHPVGLAFPLEVG---RFGQLTYVRSYQGEIKKGDITYNRTT 389
285 DTR-----TVEASEDKFTGTFVKIQANMDPKHRDRIAFMRVVSQYKGMKRLROYRT 336
390 RKVTRL-QRLARMADMMKEASTEVEVADICALRQIDAS-GDPTTEKANSGLSWESIHV 447
337 AKDVVISALTFMAGD--RSHVEAVYPGDIIIGLHNHGTTIGIDFTT---QGEWKKFTGI 390
448 PDPVISIAMKPSKNKDL--EKFSGKIQRTREDPTFKVYPTENKETIVISGGEHLLEIY 505
Qy

Db 391 PNPAPLFRRLKDPKQKQKGLKGLVQJLS-BEGAVQVFPRISSNDLIYAGVGLQEDVV 449
QY 506 AQLERREYGCPCITGKPKVA---FRETITAPVPDPFHKKQS 544
Db 450 VARLKSEYVAVESVYATARWECADAKKEEFKKNES 491

RESULT 12

US-08-984-618-8
Sequence 8, Application US/08984618
Patent No. 6251647
GENERAL INFORMATION:
APPLICANT: de Lencastre, Herminia
APPLICANT: Tomasz, Alexander
TITLE OF INVENTION: AUXILIARY GENES AND PROTEINS OF
NUMBER OF SEQUENCES: 17
TITLE OF INVENTION: METHICILLIN RESISTANT BACTERIA AND ANTAGONISTS THEREOF
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klaubert & Jackson
STREET: 411 Hackensack Avenue, 4th Floor
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/984,618
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-213 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-487-5800
TELEFAX: 201-343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 531 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: *Dichelobacter (Bacteroides) nodosus*
US-08-984-618-8

Query Match 11.3%; Score 440; DB 4; Length 531;
Best Local Similarity 25.8%; Pred. No. 1,1e-34;
Matches 143; Conservative 101; Mismatches 236; Indels 74; Gaps 18;
QY 47 RNIGISAHIDSGKTTITERVLYTGRIAKHEVKGKDG-VGAVMDSMELEROGITIQSA 105
Db 13 RTEAIIISHDAGTTITTEKLLFGCAIALAGAVKGRVAHHATSDDMKKEOBERGISVTS 72
QY 106 ATEPMKDVVNIITDPGHVDFTEVERALRVLDGAVLVCAVGVOCOTMTNRMQKRY 165
Db 73 VMOFPHGKVINLIDTPGHDFSEDYRITLTAYDSALMVIDCAKGVBERITKLMVEVCRLR 132
QY 166 NVPELFTFINKLDRMGSPARALOQMSKLNHTAFMOIPWGLEGNFKGIYDLIEBRAIYF 225
Db 133 TTFITFFVNTLDDGSEPMELIDIEIRVLIHICAPVTWPIGMGRSLKGIYHLARD-TVVF 191
QY 226 --DGDSSQIVRYE-----IPALRAAATDRHOELIECVANSDEQLGEMFLEEKIP 274
Db 192 YTTGKGASINHGELTVGLDNPRLDTLLPDIIDDFREEI-----HFLREVGNFDEH--- 243

QY 275 SISDLKAIIRATLKRSFPTVFLGSAKKNKGVQPLDAVLLEYLPNPSEVONYAILINKKD 334
Db 244 -----AYLKGELTPVFGSAINSPVEMLTDPQAQLAPPRPRRT-----TEREV 288
QY 335 SKEKTIILNSSRHNHSPVGLAFLPEVG---RFGQLYTVRSYQGLKGGDTIYNTRTYR 390
Db 289 APOEKL-----TGFFVKIQANMDLKHDRDIAFMVNSGTFRAGKKLQWVRIG 336
QY 391 KKVRL-QRLARHADMEASTEVEYVAGDICALFGIDCAS-GDTFPDKNKSGLSMESINHP 448
Db 337 REVKIPDALTFLAAREHA--QEAFAGDITGIHNGTIRIGDTYE---GSLQPTGIP 390
QY 449 D--PVISIAMKPSNNKNDLEKFSKIGRFTREDPT--FKYVPTENKETVISCMEILHEI 504
Db 391 DPAPLFRVQKDPKXKALKGLAQLCEBATOFPKRLI---GSDILGAIYVLQFEV 447
QY 505 VAQRLERREYGCPCITGKPKVAFRETITAP--VPPDFTHKKQSGAGQYGVLEPLD 561
Db 448 VQQRLETEYVAKQFESVAVATARIWEAFNDKALKQFIDKQANLHHDYEQLYIAP-- 505
QY 562 PEDYTKLERSDET 575
Db 506 --SRVNLQTOERF 517

RESULT 13

US-09-134-001C-3390
Sequence 3390, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3390
LENGTH: 634
TYPE: PR
ORGANISM: *Staphylococcus epidermidis*
US-09-134-001C-3390

Query Match 10.6%; Score 414; DB 4; Length 634;
Best Local Similarity 21.8%; Pred. No. 6e-32;
Matches 154; Conservative 91; Mismatches 202; Indels 258; Gaps 18;
QY 44 EKIRNIGISAHIDSGKTTITERVLYTGRIAKHEVKGKDG-VGAVMDSMELEROGITIQ 103
Db 25 EOVNRNIAIIAHVDHGTTLVLDLQKSGIFRENEHYDER-----AMDSNDLEREGITTL 79
QY 104 SAATFMMKDVVNIITDPGHVDFTEVERALRVLDGAVLVCAVGVOCOTMTNRMQK 163
Db 80 AKCTAIDVKGCTINILIDTPGHADFGSEVERIKMVDGVVLVDAYEGTMPQRPVLKAL 139
QY 164 RYVPELFTFINKLDRMGSPARALOQMSKLNHTAFMOIPWGLEGNFKGIYDLIEBRAI 223
Db 140 EQNLKPVVYVNNKIDPAAPR-----EGVDEVLDL----- 169
QY 224 YFDGDSQIVRYGEIPALRAAATDRHOELIECVANSDEQLGEMFLEEKIPSIDKLAI 283
Db 170 -----FIELBAN-DEQL----- 180
QY 284 RRAITLKRSFTYF-----LGSALKKKGQPLDAVLLEYLPNPSEVONYAILINKKD 334
Db 181 -----DPPVYVAVNGTASLDBSEKODENNQSIYETIIDYVAP-----VDN 222
QY 335 SKEKTIILNSSRHNHSPVGLAFLPEVGRFQLYTVRSYQGLKGGDTIYNTRTKAYR 394

```

Db      223 SBEPLQFQALADYNY-----VGRIG--VGRVFRGMRGVADVSLKLDGTAK 269
Qy      395 LORLAM--HADMEASTEVEVYAGDICALFGI--DCASGDTPTDKANSGLSMESIHVPDV 451
Db      270 NRRVTKIFEGYFGKREIEEAOAGDLIANSMDIVGETVTPHHRD--PLPLRLIDEPT 328
Qy      452 ISIAMKPSKN-----DLEKFSKIGRTREDPTFKVYFDTEKTEVIISGMELH 502
Db      329 LEWTFKVNNSPFAREGDYVTAQIOERLDQLETVSLKVTPTDQPSGVVAVAGRELUH 388
Qy      503 ELYAOLREYGCPCITGKPKVAFRETTAPVPFDTHKKQSGAGQYGVIGVLEPLD 562
Db      389 SLIEMWRE--GELOVSKRQVILRE-----IDGVLS----- 419
Qy      563 EDYTLKFSDETFGSGNIPKQFVPAVEKFLDACCEKPLSGHKLGLRFLVLDGAAHMDVS 622
Db      420 ----- 419
Qy      623 NEISFIRAGEGALKOALANATLCILEPIMAVEVAPNEROGVYAGINRRHG-----VITG 678
Db      420 -----EPFERVQCEVSENAAGVIESLGARKGEMLDMMTT 454
Qy      679 QGVEDYFTLYADVPLNDMFGVSTELRSCTECKGERTTMEYSRQP 723
Db      455 DNGLT---RLIFWVPARGMIGYTTETEMSWTRGYGIINHTFEERFP 496

```

RESULT 14 US-09-134-001C-3574

```

/ Sequence 3574, Application US/09134001C
/ Patent No. 6380370
/ GENERAL INFORMATION:
/ APPLICANT: Lynn Doucette-Stamm et al
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
/ TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
/ FILE REFERENCE: GTC-007
/ CURRENT APPLICATION NUMBER: US/09/134,001C
/ PRIOR FILING DATE: 1998-08-13
/ PRIOR APPLICATION NUMBER: US 60/064,964
/ PRIOR FILING DATE: 1997-11-08
/ PRIOR APPLICATION NUMBER: US 60/055,779
/ PRIOR FILING DATE: 1997-08-14
/ NUMBER OF SEQ ID NOS: 5674
/ SEQ ID NO 3574
/ LENGTH: 531
/ TYPE: PRF
/ ORGANISM: Staphylococcus epidermidis
/ US-09-134-001C-3574

```

Query Match 9.4%; Score 367, DB 4, Length 531,
Best Local Similarity 20.3%, Pred. No. 2e-27,
Matches 137, Conservative 99, Mismatches 192, Indels 248, Gaps 19,

```

Qy      44 EKIRNIGISAHIDSGKTTLTERLVYYTGRIAKMHEVKGKCGVAGVADSMELERQGITIQ 103
Db      45 ENIRNFSIAIHDGKSTLADRILENT-----KSVETREMQDQLLSDMLEBERGITIK 98
Qy      104 SAATPTMM--KD---VNIIIDTPGVADTTIEVERLARLYDGAVALYLCVAGVQCCQCTMTV 158
Db      99 LNAVRLKYLAKGGETYTFHLIDPGVADTYEVSRLACEGAILVDAOGIEAQTLAN 158
Qy      159 NROMKRYVPLTFINKLDRMGSNPARALQOMRSKLNHTAFQIQPMGLEGNFKGIVLDI 218
Db      159 VTLALDNDIELLPVANKIDLPAAEPKVQOE----- 189
Qy      219 EERRAIVFDIDFSGQIVRYGEPALRLAAATDHRQELIECVANSDEQVGENFLEKIPISID 278
Db      190 -----LEDVIGIDQE----- 199
Qy      279 LKLAIRRALTKSFTTPVFLGSALKNQVOPLDVAULVEYFNSEVQVYAILKKQDSKEX 338
Db      200 -----DVLASAKSNIGIEILLEKIYDVVFAPD-----GDPEAP 233

```

```

Qy      339 TKILNSSRHNSHPVGLAPLEVGFRFGQITVRSYOGELKIGDTIYNTTRKKVYLRQL 398
Db      234 LKALIFDSEYD--PYRGV-----ISSIRIIDGVVYKAGDHIKMMATGEFEVTEV 280
Qy      399 APMHADMEASTEVEVYAGDICALFGI-----DCASGDTPT-----DKANSGLSMESIHV 447
Db      281 GINTPKOL--PYEEILTVGCVIILASIKVNDSDRVGDTITTLAERPADKPLQGYKK----- 333
Qy      448 PDPVISIAMKPSKNNDLEKFSKIGRTREDPTFKVYFDTEKTEVIISGMELH 502
Db      334 MRPWFPGIFPIDNDKVDNDRBALKQLNDLSLE--FEPSSQALGFRYRTGFIGMLHM 391
Qy      503 ELYAOLREYGCPCITGKPKVAFRETTAPVPFDTHKKQSGAGQYGVIGVLEPLD 562
Db      392 EIQRIERIEFEIILATAPSVIYQ----- 416
Qy      563 EDYTLKFSDETFGSGNIPKQFVPAVEKFLDACCEKPLSGHKLGLRFLVLDGAAHMDVS 622
Db      417 -----C-----ILKDG----- 422
Qy      623 NEISFIRAGEGALKOALANATLCILEPIMAVEVAPNEROGVYAGINRRHGVTGQGV 682
Db      423 SEVSVDNPAQMERKXIEH---IYEPVKATMVPNDVGAVMELCQRKRGQFTIMDYL 478
Qy      683 EDY--FTLYADVPLNDM 697
Db      479 DDIRVNIHETPLSEV 494

```

RESULT 15 US-09-095-855-207

/ Sequence 207, Application US/09095855
/ Patent No. 6160083

```

/ GENERAL INFORMATION:
/ APPLICANT: Tan, Paul
/ APPLICANT: Visser, Elizabeth
/ APPLICANT: Skinner, Margot
/ APPLICANT: Prestidge, Rose
/ TITLE OF INVENTION: Compounds and Methods for
/ TITLE OF INVENTION: Treatment and Diagnosis of Mycobacterial Infections
/ NUMBER OF SEQUENCES: 208
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: Law Offices of Ann W. Speckman
/ STREET: 2601 Elliott Avenue, Suite 4185
/ CITY: Seattle
/ STATE: WA
/ COUNTRY: USA

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/095,855
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/705,347
FILING DATE: 29-AUG-1996
APPLICATION NUMBER: 08/873,970
FILING DATE: 12-JUN-1997
APPLICATION NUMBER: 08/997,362
FILING DATE: 23-DEC-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sleach, Janet
REGISTRATION NUMBER: 37,007
REFERENCE/DOCKET NUMBER: 11000.1002c3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-269-0565
TELEFAX: 206-269-0563
TELEX:

/ INFORMATION FOR SEQ ID NO: 207:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 173 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-09-095-855-207

Query Match 8.7%; Score 338.5; DB 4; Length 173;
Best Local Similarity 39.7%; Pred. No. 1.9e-25;
Matches 69; Conservative 37; Mismatches 65; Indels 3; Gaps 1;

QY	569	EFSDFTGNSNIPKQFVPAVEKGFIDACEKGPLSGHKLGLRFVLDGAAHHMVDNSNISFI	628
DB	3	EFENKVTGGRIPREYIPSVNAGAQDAMQYGLAGYPLVNVKLTLLDGAHYEVDSSSEMAFK	62
QY	629	RAGEGALKQALANATLCLILEPMAVEVAPNEFQGOVIAGINRRHGVTGQDGVEDYFTL	688
DB	63	VAGSQVMKKAAAOAPVILEPVMAVEVETPEDYMGVEYIGDLSRRGQIQMEERSGARVY	122
QY	689	YADVPLNDMGYSTELRSCTEGEGEYTMESRYQPCLPSTQEDVINKYLEATGQ	742
DB	123	KAQVPLESMFGYVDLRSKTQGRANYSMPDSTAEVPAVNSKEIIAK--ATGQ	173

Search completed: July 14, 2003, 18:24:35
Job time : 12.4425 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:15:27 ; Search time 17.6298 seconds
(without alignments)
4965.877 Million cell updates/sec

Title: US-09-815-379-6

Perfect score: 3898
Sequence: 1 MRLGAAVAALGRGRAPAS.....INKYLEATGQLPKKAKAKN 752

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications AA.*
2: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUECOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUECOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUECOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US05_PUECOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUECOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUECOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3898	100.0	752	9	US-09-815-379-6
2	1489	38.2	692	10	US-09-815-242-11568
3	1454	37.3	700	10	US-09-815-242-11060
4	1427	36.6	704	10	US-09-913-020-248
5	1427	36.6	704	10	US-09-815-242-10363
6	1412	36.2	705	9	US-09-738-626-4055
7	1409.5	36.2	709	9	US-10-156-761-12453
8	1406	36.1	702	10	US-09-815-242-11858
9	1374	35.2	704	10	US-09-815-242-13965
10	1352.5	34.7	693	10	US-09-815-242-4481
11	1352.5	34.7	693	10	US-09-815-242-10481
12	1352.5	34.7	693	10	US-09-815-242-13233
13	1351.5	34.7	715	10	US-09-815-242-12443
14	1351.5	34.7	715	10	US-09-815-242-12736
15	1338	34.3	693	10	US-09-815-242-5238
16	1173	30.1	696	9	US-10-156-761-9238
17	667	17.1	732	9	US-10-156-761-14353
18	543.5	13.9	663	9	US-10-156-761-15035
19	516	13.2	844	9	US-10-108-605-185

20	505	13.0	845	9	US-09-991-496-110	Sequence 110, App
21	505	13.0	845	10	US-09-874-923-110	Sequence 110, App
22	478.5	12.3	884	9	US-10-128-714-4037	Sequence 4037, App
23	474.5	12.2	520	10	US-09-815-242-5735	Sequence 5735, App
24	436.5	11.2	599	10	US-09-815-242-11326	Sequence 11326, App
25	417.5	10.7	1087	9	US-10-128-714-8410	Sequence 8410, App
26	417	10.7	1013	9	US-10-128-714-3410	Sequence 3410, App
27	416	10.7	601	10	US-09-815-242-12358	Sequence 12358, App
28	416	10.7	604	10	US-09-815-242-5260	Sequence 5260, App
29	401.5	10.3	591	10	US-09-815-242-10419	Sequence 10419, App
30	396	10.2	557	10	US-09-815-242-13521	Sequence 13521, App
31	393.5	10.1	549	10	US-09-738-626-4536	Sequence 4536, App
32	393.5	10.1	607	10	US-09-815-242-13791	Sequence 13791, App
33	392.5	10.1	516	10	US-09-815-242-11136	Sequence 11136, App
34	387.5	9.9	632	10	US-09-815-242-11727	Sequence 11727, App
35	383	9.8	602	10	US-09-841-132-495	Sequence 495, App
36	381	9.8	602	10	US-09-841-132-565	Sequence 565, App
37	377	9.7	622	9	US-10-156-761-13093	Sequence 13093, App
38	368	9.4	635	9	US-10-156-761-10714	Sequence 10714, App
39	361.5	9.3	602	10	US-09-815-242-10802	Sequence 10802, App
40	353	9.1	605	10	US-09-815-242-12087	Sequence 12087, App
41	349	9.0	637	9	US-09-738-626-4721	Sequence 4721, App
42	338.5	8.7	173	9	US-10-051-643-207	Sequence 207, App
43	327	8.4	615	9	US-09-738-626-6067	Sequence 6067, App
44	325	7.1	144	9	US-09-895-913A-156	Sequence 156, App
45	217	5.6	408	10	US-09-845-335-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-09-815-379-6
; Sequence 6, Appl1 US/09815379
; Publication No. US20030073613A1

GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS
; TITLE OR INVENTION: ENCODING THE SAME
; FILE REFERENCE: 10716/35
; CURRENT APPLICATION NUMBER: US/09/815.379
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 60/191,134
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 6
; LENGTH: 752
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-379-6

Query Match	Score	Length	DB	Length	DB	Score	Length	DB	Length	DB
Best Local Similarity	100.0%	752	0	Mismatches	0	Indels	0	Gaps	0	
Matches	752	Conservative	0	Mismatches	0	Indels	0	Gaps	0	
1	MRLGAAVAALGRGRAPASLGMORQVNMWACRWSGVLPEKINIGISAHIDSCT	60								
61	TTTEVLYTGTIAGHIEVKGADGVAVMDSMELEBROGITQSAATFTMMVDVINIID	120								
61	TTTEVLYTGTIAGHIEVKGADGVAVMDSMELEBROGITQSAATFTMMVDVINIID	120								
121	TPGHVDFTEVERALRVLDGAVLYICAVGVQCQTMTVNRQKRYNVPFLTFINKLRNG	180								
121	TPGHVDFTEVERALRVLDGAVLYICAVGVQCQTMTVNRQKRYNVPFLTFINKLRNG	180								
181	SNPAAALQOMSKLNHTNAPFQIIPMGLEGNKRGIVDLIEBAIYFDGSPQIVRGELPA	240								
181	SNPAAALQOMSKLNHTNAPFQIIPMGLEGNKRGIVDLIEBAIYFDGSPQIVRGELPA	240								

```
QY 241 ELRAAATDHRQELIECVANSDEOLGEMFLEEKIPISIDLKAIARRATLKRSFTPVFLGSA 300
DB 241 ELRAAATDHRQELIECVANSDEOLGEMFLEEKIPISIDLKAIARRATLKRSFTPVFLGSA 300
QY 301 LKNKGVPQLDVLAVLELVPNSEVQNTAILNKODSEKTKILMNSSRRHNSHPVGLAPPL 360
DB 301 LKNKGVPQLDVLAVLELVPNSEVQNTAILNKODSEKTKILMNSSRRHNSHPVGLAPPL 360
QY 361 EVGRFQGLTVVRSYQOSELKKGDITTYNTRTKKTVRLRLARMHADMMEASTEEVYAGDICA 420
DB 361 EVGRFQGLTVVRSYQOSELKKGDITTYNTRTKKTVRLRLARMHADMMEASTEEVYAGDICA 420
QY 421 LFGIDCASGDTFTDKANSGLSMESIHPDPVISANKPNKNDLEKRSKIGRFTREDDT 480
DB 421 LFGIDCASGDTFTDKANSGLSMESIHPDPVISANKPNKNDLEKRSKIGRFTREDDT 480
QY 481 FKTYFTENKETIYISGGEHLHEIYAQRLEBYGCGCTIGKPKVAARETTTAAVPDPFTH 540
DB 481 FKTYFTENKETIYISGGEHLHEIYAQRLEBYGCGCTIGKPKVAARETTTAAVPDPFTH 540
QY 541 KKGSGAGGVGKVIYGLPELPDEPDYTKLEFSDTFCPSNIPKOPVPAVEKGFLLDACEKGPL 600
DB 541 KKGSGAGGVGKVIYGLPELPDEPDYTKLEFSDTFCPSNIPKOPVPAVEKGFLLDACEKGPL 600
QY 601 SGHKLSGRLREVLDDGAHHWDSNEISFIRAGEBALKQALANATLCLEPIMAVEVVAAPNE 660
DB 601 SGHKLSGRLREVLDDGAHHWDSNEISFIRAGEBALKQALANATLCLEPIMAVEVVAAPNE 660
QY 661 FQGVYAGINRRHGVITGQDGVEDYFTLVADVPLNDMFGYSTLRCTBGKGEYTMESYR 720
DB 661 FQGVYAGINRRHGVITGQDGVEDYFTLVADVPLNDMFGYSTLRCTBGKGEYTMESYR 720
QY 721 YQCLPSTQEDVINKYLEATGOLPVKKGAKAN 752
DB 721 YQCLPSTQEDVINKYLEATGOLPVKKGAKAN 752

RESULT 2
US-09-815-242-11568
; Sequence 11568, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haebebeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: PaacSeq for Windows Version 4.0
; SEQ ID NO 11568
; LENGTH: 692
```

```
; TYPE: PRF
; ORGANISM: Helicobacter pylori
US-09-815-242-11568

Query Match      38.2%; Score 1489; DB 10; Length 692;
Best Local Similarity 45.7%; Pred. No. 1.3e-115;
Matches 319; Conservative 128; Mismatches 233; Indels 18; Gaps 11;

QY 42 PNEKIRNIGISAHISGKTTTLTERLVLYTGTGRJAKKHEVYKQGVANVDSMELEROGIT 101
DB 6 PLNRIRNIGIAHIDAGKTTTSERLFTTGVSHKIGEV--HDG-AATYDMMEQEKERTT 62
QY 102 ICSAATFTWMKDVNINIIDTPGHVDFTEVERBALVDGAVLVLCVAGVQCQTMTVNRQ 161
DB 63 ILSAATFTWMKDVNINIIDTPGHVDFTEVERBALVDGAVLVLCVAGVQCQTMTVNRQ 122
QY 162 MKRYVPPFLTFINKLDRKGSNPARLLOQMSKLNNTAPMOJPMLEGNFKIVULLIER 221
DB 123 ANKYGVPRIVFVNRKDRIGANFYVNEQIKORLKNPVPINIPGAEDTFIGVIDLVQMK 182
QY 222 AIFYGDFPSQIVRYG--SIPAEELRAATDHRQELIECVANSDEOLGEMFLEEKIPISIDL 279
DB 183 AIVWNNB--TMGAKTYVEEIPSDLEKAKOYREKLEVAABQDLEMEKTLGSEELDIEBI 241
QY 280 KLAIRATLKRSFTFVLGSLAKNKGVQPLDVLAVLELVPNSEVQNTAILNKODSEKKT 339
DB 242 KGIKTGCLNMSFVWMLGSSSFKNKGVTLLDAVIDYLPAPFEVVDI---KGIDPKTEE 297
QY 340 KILMNSSRRHNSHPVGLAFLPEVGRF--GQITVRSYQOSELKKGDITTYNTRTKKTVRLQL 398
DB 298 EVFVVS--DDEEFAGLAFKIMTDPFVQQLTFVRVYRGLESGSYVYNSYTKDKKERVRL 355
QY 399 AMHADMMEASTEEVYAGDICALFGI--DCASGDTFTDKANSGLSMESIHPDPVISANK 457
DB 356 LKMSNKE--DIKEYIYAGEICAFVGLKDTLTJDTLDCDKNA--VUERNEFPFPVHIAYE 413
QY 458 PSNKNDLEKFSKIGRFTREDDTFFKYVFTENKETIYISGGEHLHEIYAQRLEBYGCGC 517
DB 414 PKTKADQEKMGVALDKLABEDPSFRVMTQEEVGTILGSGMELHEIYIDRLKREFKVA 473
QY 518 ITGKRVAFRETTTAAVPDPFTHKKQSGAGGVGKVIYGLPELPDEPDYTKLEFSDTREGS 577
DB 474 EIGQPOVAFRETTIRSVSEKHKYAKQSGRGYGHVFTKLEKPEG--SGYEFVNEISG 531
QY 578 NIPKOPVPAVEKGFLLDACEKGPLSGHKLSGRLREVLDDGAHHWDSNEISFIRAGEBALK 637
DB 532 VIPKEIIPAVDGIQDEANQVLAGYPVVDKVTLYDSYHVDVDSSENAFKIAGSMATFE 591
QY 638 ALANATLCLEPIMAVEVVAAPNEFQGVYAGINRRHGVITGQDGVEDYFTLVADVPLNDM 697
DB 592 ASRAANPVLLEPMKVEVEVPEEYMGDIVGLNRRRQINSMDRLGLKIVNAFVPLVEM 651
QY 698 FGYSTELRSCTBGKGEYTMESYRYPCLPSTQEDVINK 735
DB 652 FGYSTELRSATQGRGTYSMEDPHYGEVPSNIAKEIYEV 689

RESULT 3
US-09-815-242-11060
; Sequence 11060, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haebebeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
```



```

Dh KVNIGIMHIDAGKTTTTERILFYGINRKXGFE--HDG-GATTDMDEGKEKGIITTS 63
Qy 105 AATFTMKKDVNNINIITDPGHVDFTTIEVERALVDLGAVLVLCAVGVQOQMTVNRQMKR 164
Dh 64 AAATVCEDNNQVNIIDITDPGHVDFTTIEVERSLRVLGAAVAFPGCKGVEYPOSEQWROATK 123
Qy 165 YNVPFLTFLNKLDRBMSNPBARALQCRSKLNHNTAFMOJFMLEBENFGIYDLLEBAIY 22
Dh 124 YDVPRLCEFNKMDKLADFPYFTVTGTTIEDLGAKPLWMOJFJGAENPFGVIDLLEMKALT 183
Qy 225 FDDDFSQIYVYG-----EIPAEIARAADNRHOELIECVANSDEOLGEMFLEEKIPIISD 278
Dh 184 WRG-----VPIGTEATYEVEIPIALBARABEYKLEIYAEDEDEIMARKYGGESLTAIE 233
Qy 279 LKLAIRRAATLKRSFTFVFLGSAIKNKGVOPLLDAVLEYLFENSEYONAIINKKDSKEK 338
Dh 240 IKAIRRMVNVSEIPIYVCGTAYKNGKIQPLDAVDFLESPLDG---TKGDYKDP 295
Qy 339 TKILMNSIHNHNPFGVLAAPLEVEGR-FQOLTYVASYOGELKKGTTIYNTTRCKKVALQR 397
Dh 296 EKVYLRKP-SDEEPLSALAFKJAAHPFCKLTFEVLYSQVYEPGQVLYNSTKNKERIGK 354
Qy 398 LARWADMEASTEEBYAGDICALFGI-DCASGDPFTDKANSGLMESIHVDPVISIAM 456
Dh 355 LFGOMHAR-KENPVEVAHAGNIYAFIGLKTDTTGDLCD-AANPILMESMDPPYIOVAI 412
Qy 457 KPSNKNDLKFKSGIGRFTREDPTFKVYEDTENKSTVJSGMELHLEIYAQRLEREXGCP 516
Dh 413 EPKTKSDQKLGVAIKQLABEDPTFVHLDDSGQTVIGCMGELHLDLVLRMRKREFVE 472
Qy 517 CITGKRVNAFRETTITAPV-PFDETHKKGSGAGQYGRKIVGLEPIEDP-----DYTKL 566
Dh 473 ANIGDPOVNAFREIRKRVSLSTYHKKQYGGSGQPAKVIJITEPYAPADELEBEGEAIY 532
Qy 569 EFSDETFGSNIPOPFAYAVEKGLDCEKGPISGKHLSGLRVLDDGHHWVDSNEISFI 628
Dh 533 KENAVATGSRVPERYIPSDAGIQDMQOGFLAGIPLVNVKATLEDGAYHVDSSERAFK 592
Qy 629 RAGEGALKQALANATLCTLEPIMAVEVVAENFEGOVLAGINRRHGVITGQGVEDVFTL 688
Dh 593 LAGSQAPKVAKAAKAVLLEPIHNSVEITTPPEYMEBEIVGNVNSRGGQJASMDDBAGKLV 652
Qy 689 YADVPLNDMEFGYSTELRSCTEGKGEYTMESRYQPCLPSTQEDVI 733
Dh 653 KAKVPLSQMFGYVGDLSRKTQGRANYSWTFSDYAEVAPNAVADVI 697

RESULT 7
US-10-156-761-12453
Sequence 12453, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMIURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156, 761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 12453
LENGTH: 709
TYPE: PRF
ORGANISM: Streptomyces avermitilis
US-10-156-761-12453

```

```

Query Match 36.2%; Score 1409.5; DB 9; Length 709;
Best Local Similarity 43.6%; Pred. No. 5.8e-109;
Matches 308; Conservative 126; Mismatches 247; Indels 25; Gaps 11

QY KIRNIGISAHIDSGKTTLTERRVLYYGRIAIOMHEVKGKQGVAVWDSMELRQRGRTTQS 104
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db KVRNIGIMAHIDAGKTTTERTTLTFYGVSYXIGEY--HDG-AATMDMKQEGERGRTTIS 66
105 AATTMM---KDVNINIIDTPGHVDFITIEVERALRVLDGAVLVLCAGVQOCQMTYNR 160
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db AATTCHEPLEDNDDYTIINIIDTPGHVDFITIEVERSLRVLDGAVTVDFGAVGVEPOSETWNR 126
161 QMKRYNVPLFEINKLIDMGSNPAPALQOMSKLNHNTAFPOICPMGLGNPFGIYDLTEE 220
127 QADRGVPRICEVKNLIDRTGAEFHRCVMTIKORIGAVTIIQOLPGALEMDFGVDVLVM 186
QY RAIYFDGDSQIVRYG--EIRPAELRAAATDHRQBLIECVANSDBQLGEMFLEBKIPISID 278
Db KALVWSAAKAGENVYDVVIDPATHAABAEYRKLVEFVAENDEIDEMFLFEGQPTREQ 246
187
QY LKAIIRRTLR-----SFTPVPLGSLKKKKQVPLLDVLEIYIPNSSEVQNTAILNKK 332
219 LKAIIRRTLR-----SFTPVPLGSLKKKKQVPLLDVLEIYIPNSSEVQNTAILNKK 332
247 LYAIIIRRTTISAGKGDGVTVTPFGCITFKKKGVQPLLDVAVRYLPPLDVEAIEGHQVK 306
QY DDSKERTIILNNSRHNHPFVGLAF--LEVGRFQGLYYVRSYOGELKKGDTIYNTPTRK 391
Db D-----PEVVKKRPSDEBEPISALNFKTMSDPHLKLTFFRVYSGRLESSTAVLNSVKK 361
307
QY KVRLORLARMAHADMMAEASTEEVYAGDICALFGI--DCASGDTFTDKANGLSMESIHPDP 450
392
Db KERIGCIYRMANKRE--EIEAVGAGDIYAVVMGLKQTTGETLSDCKNP--VILESMDPAP 419
362
QY VISIAMKSNQNDLEKSKGIGRFPREBPTKRVYPTDENKEFVISGNEIHLIETAOBLE 510
451
Db VIOVAIEPKSGKDEKGVAIORLAEBDPSFOVHSDERTGOTIIGGNEIHLIETAVDRKR 479
420
QY REYGCPTCTPKPAFRETITAPVP--PDTKKKQSGAGQYKVGIVLEPLDPEDTYKLE 569
511
Db REFKEYANVGFPQVAYRRTTIKTYERVDYTHKKQGTGCGQPAKVQIAIIEPTETGPAYE 539
480
QY FSDETFGSNIDPKQFPAVEKGLFDLACEXKPLSGHKLGLRVLPDQGAHHMVDNSIESPR 629
570
Db FVNKVTGGRIIDREYIIPSDVAGAQBAMQGLIAGYEMTGVRVTLIDGAYHVEDSSELAFKI 599
540
QY AGEBAIKQALNATLTCTIEPIIMAVEVYAPNEFQGVYIAGINRRHGVITYGQGVEDYFTLY 689
630
Db AGSQAFKAAKAPSVLLEPMMAVEVTPPEESMGVVIDLNSRRQIQIAMEERSGARVVK 659
600
QY ADVPLDMFGSTELRSTCEKGEXTMYSYQPCLEPSTQRBVDINK 735
690
Db GLVPLEMEFGYVDLRKTSGRASYSQFDSYAEVPRNVAEETIAK 705
660

RESULT 8
US-09-815-242-11858
; Sequence 11858, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zykkind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: EITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815.242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

```

```

/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 11858
/ LENGTH: 702
/ TYPE: PRT
/ ORGANISM: Pseudomonas aeruginosa
/ US-09-815-242-11858

```

Query Match 36.1%; Score 1406; DB 10; Length 702;

Best Local Similarity 43.6%; Pred. No. 1.1e-108; Indels 30; Gaps 13;

Matches 308; Conservative 121; Mismatches 248;

```

QY 42 PNEKINIGISAHIDSGKTTLTERVLYTGRIAKMEHVKGKDGAVGAVMSMELERQGIT 101
DB 6 PIELYINIGIVAVDAGKTTTERILFTGVNKHGEV--HDS-AATMDMVEQGERGIT 62
QY 102 IGSAAFTWKKD-----VNIIITDPGHVDTTIEVERALVLDGAVLVCAVGVOC 153
DB 63 ILSAATTAFTWQSGTKOPPHRYRNIIIDTDPGHVDTTIEVERSLVLDGAVVFGAGGVGP 122
QY 154 QMTTVAROMKRYVVPFLTFINKLDRMGSNPARALQOMRSKLNNTAFMOIPMGLGNFPG 213
DB 123 QSTTVROANKIVPEPLAVNKKRQADFLRVAAQIKQLGHPVPPIQIATIGSENFSG 182
QY 214 IVDLIERAIYF-DGDFSQIVRYGEIPABLRRAATDHRQELIECVANSSEQLGEMPLEEK 272
DB 183 QIDLVGMKAIYVMDADQGSYREBEIPALRALAEEMRAHMAEADELNNKYLEEB 242
QY 273 IFSISDLKAIIRATLKRSFTPVFSGALKKKGVQPLDAVLEYLNPSEVQNYAIALNK 332
DB 243 ELISIEIKKGLRQRTLANQIVPAVLGSSFKNKGVPLVDAVIDYLPAPSEIP--AIRGTD 300
QY 333 DSKERTKILMNSRH--NSHPFVGLAFPLEVGRF-GOLTVYRSYSGELKKGDTIYNTRT 389
DB 301 PDDEEK-----HERRAADDEPFSALAFKIATDPFVGILTFARVYSGVLTSGLAVLNSVK 355
QY 390 RKKVRLQRLARMAHMMEAESTEVEVAGDICALFGI--DCASGDTF--TDKANSGLSMESI 446
DB 356 GKKEKRGVAVQWANGRD--EIKEVRAGDIALALGMQDVTGDTLCALDKP---IILERND 411
QY 447 VDDPVVISIAMPKSNKNDLEKFSKIGRFTREDPTFYVYPTENKETVIGSGMLHEIYA 506
DB 412 FPDPIVSIVAVEPKTKADQEKMGIALSKLAQEDPSFVKTDDEFAQTIIISGMGELHDIIV 471
QY 507 ORLERVYGCPCITGKRVAPREFITAPVPEDFTHKKOSGAGQGVAVIGVLEPLDEEDYT 566
DB 472 DRRRRRFGVANTGKQVAVRETIRNTCEIEGKFFVQSGGRGFGHCWIRFAPAD--EGGE 530
QY 567 KLEFSDTEGTSNIPKQFVPAVEKGLDACEKGPLSGHKLSGLRFVLQDGAHNVADNEIS 626
DB 531 GLEFHNIEVGVAPREFITAIQIGIEDQNGVLAQYPLIGLKATYVDSGSHYVDSSEWA 590
QY 627 FTRAGGALKQALANATLCTLEPIYAVEVVAENEFQGVYAGINRHHGVITGQDGVEDYF 686
DB 591 FKLAASMATKQLSQKGAVALLEBVMKVEVYTPEDYGDVWGDNRRRRGLIQGMEETPAKG 650
QY 687 TLYADVPLNDMPGYSITELRSCTEGKGEYMEYRQPCLPSTQEDVI 733
DB 651 VTRAEVPLGEMFGYATDVRSKQGRASYMEFVRYAEVSPASVAGIV 697

```

RESULT 9

US-09-815-242-13965

Sequence 13965, Application US/09815242

Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert

APPLICANT: Olsen, Karl L.

APPLICANT: Zyskind, Judith W.

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John D.

APPLICANT: Carr, Grant J.

APPLICANT: Yamamoto, Robert T.

APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in

TITLE OF INVENTION: Prokaryotes

FILE REFERENCE: ELITRA.011A

CURRENT APPLICATION NUMBER: US/09/815,242

CURRENT FILING DATE: 2001-03-21

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 13965

LENGTH: 704

TYPE: PRT

ORGANISM: Salmonella typhi

FEATURE:

NAME/KEY: VARIANT

LOCATION: (1)...(704)

OTHER INFORMATION: Xaa = Any Amino Acid

US-09-815-242-13965

Query Match 35.2%; Score 1374; DB 10; Length 704;

Best Local Similarity 42.6%; Pred. No. 5.3e-106; Indels 36; Gaps 15;

Matches 306; Conservative 126; Mismatches 250;

```

QY 42 PNEKINIGISAHIDSGKTTLTERVLYTGR--RIAKMEHVKGKDGAVGAVMSMELEROR 98
DB 6 PIARRNIGISAHIDAGKTTTERILFTGVNKHIGEXH----DG-AATMDMVEQGER 59
QY 99 GTTISAAFTWKKV-----NIIITDPGHVDTTIEVERALVLDGAVLVCAVGV 151
DB 60 GTTISAAFTWQSGTKOPPHRYRNIIIDTDPGHVDTTIEVERSKVLELDGAVVYGAAGV 119
QY 152 QOQMTTVAROMKRYVVPFLTFINKLDRMGSNPARALQOMRSKLNNTAFMOIPMGLGNF 211
DB 120 QOQSTTVROANKIVPEPLAVNKKRQADFLRVAAQIKQLGHPVPPIQIATIGSENFSG 179
QY 212 KGIIVDLIERAI-YDGFPSQIVRYGEIPABLRRAATDHRQELIECVANSSEQLGEMPLE 270
DB 180 TGVVDLVGMKAIINMDADQGVTFEYEDIPADWQDLANEHQVLISAAASELMKYLEG 239
QY 271 EKIPIISDLKAIIRATLKRSFTPVFSGALKKKGVQPLDAVLEYLNPSEVQNYAIALN 330
DB 240 GELTEEBEIKQALRQRTLANQIVPAVLGSSFKNKGVPLVDAVIDYLPAPSEIP--AIRGTD 297
QY 331 KDDSKERTKILMNSRHNS--HPFVGLAFPLEVGRF-GOLTVYRSYSGELKKGDTIYNT 387
DB 298 IIDDGKDTF-----AERRHSDDEPFSALAFKIATDPFVGILTFARVYSGVLTSGLAVLNS 352

```


QY 388 RTKRYRLCLRLARMADMEASTEVEYAGDICALFGI-DCASGDTFTDKANSGLSMESI 446
DB 353 VTAARERFORIVQMANKRE-EIKEVRAGDIAAIGLKQVTTGDTLCD-PEXPILLERME 410
QY 447 VDDPIVISTAMKSNKNDLEKFSKICGRFRREDPTPRVYDPTENKEVISMGEHLLEYA 506
DB 411 LPEPIVISAHEBKTADXXKMGALGRLKEDPSFPVWDEESNQTIIAGMELHLDIIV 470
QY 507 ORLEREYGCPCITGKPKVAFRETITAPV-PFDTHKKQSGAGQYQKVIIGVLEPLDEPY 565
DB 471 DMKKEFNVANVGRQVAYREAIRAKYTDIDEGKAKQSGGRQYGHVIDYVLEPESGN 530
QY 566 TK-LEPSDETFGSNIPKQFVAVNEKGFUDACEKPLSGHLSGLRVLQDGAHNVDSNE 624
DB 531 PKGYEFINDIKGVIFGEYIIPANDKGIQGLKSGPLAGYPVVDLGVRLHFGSYHDVDSSE 590
QY 625 IFFIRAGEBALQKALANATLCLLEPIMAVEVAVNPFQGVAGINRRRGVITGDDGVND 684
DB 591 IAFKLAASIAFKEGFKAKPVLLEPIMKVEETPEENTDVGIDLSRRXMKLGQSESVT 650
QY 685 YFTLYADVPLNDMFGYSTELASCTEGKEVTMEYSRYQCLPSTQEDVINKYLEATGQ 742
DB 651 GKRIHAEVYLMKMFCAKXGRSXTGRASVTMKFLKYD---XAPXVAVIEMARGK 704

RESULT 10
US-09-815-242-4977
Sequence 4977, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4977
LENGTH: 691
TYPE: PRF
ORGANISM: Enterococcus faecalis
US-09-815-242-4977

Query Match 34.7%, Score 1352.5, DB 10, Length 691;
Best Local Similarity 42.1%, Pred. No. 3, 2e-104;
Matches 292, Conservative 111, Mismatches 256, Indels 15, Gaps 9;
QY 44 EKIRNIGISAHIDSGTTLTERLYTGRARAKHEVKGSDVAVNDSMELEGRITITQ 103
DB 8 EKIRNIGIMAHVAGKTTTTERLYTGRARAKHEVKGSDVAVNDSMELEGRITIT 64

QY 104 SAATFTMKNDVNNIITDPGHVDTIEVERALRVLDGAVLVLCAGVQOCOTMYNRQMK 163
DB 65 SAATTAQMKGYRVNIITDPGHVDTIEVERALRVLDGAVLVLCAGVQOCOTMYNRQMK 124
QY 164 RYNNVEFLFIKLDMSGNPARALOQMSKLNHTNAPFOIMGEGNKGIVDLLEBAI 223
DB 125 EKVPRIVFCNMKMDIGADPFYSVESLHRLQANNAHPQIPIGAEEDTGTIIDLKQAB 184
QY 224 YFDGDSQIVRYGEIPALRAAATDHPQELIECVANSDEQLEMPLEEKISIDLKAI 283
DB 165 IYTNLGDCTIDQETDIPEDYLEKAQEMREKLYEVAVAETDEDLMMKYLEEBEILVAGI 244
QY 284 RRATLKRSFPVFLGSLAKNGVQPLDLAVLEBYLPNPSVQNYALINKKDSKEKTKILM 343
DB 245 RQATINVEFPVLGASAFKNGVQLMDAVLDVLPSPIDAIKIDIKYTD-EETTR--- 300
QY 344 NSSRNSHPFVGLAPPLEVGRF-GQLTYVRSYQGLKKGDTIYNTRTKRYRLQRLAMH 402
DB 301 --PADDEAPFASLAKVMTDPPVGRITFRFVYSGVLESQYVLNASKKKEKRIHILQMH 358
QY 403 ADMMEASTEVEYAGDICALFGI-DCASGDTFTDKANSGLSMESIHPDPVISIANKPSNK 461
DB 359 ANTRQ-ELDKYSSGDIILAAVGLKQTTTDTLC-ALDAPVILLESIEFPDPVIGVAVRPSK 416
QY 462 NDLEKFSKIGRFTEDPTFKVYFDTENKEVISMGEHLLEYAORLEREYGCPCITGK 521
DB 417 ADQDQGVVALQGLAEEDSPFRVETNVEETGEFVISMGEHLHLVDLRMKREFKVBANYGA 476
QY 522 PKVAFRETITAPVPPDPTFHKKQSGAGQYQKVIIGVLEPLDEPDYTKLEFSDETFSSNIPK 581
DB 477 PQVSRERETFRATKKEGKRVKQSGQYGHVWVEFTP--NEEVKGFEPENAVIGGVPR 534
QY 582 QFVPAVEKGFUDACEKPLSGHLSGLRVLQDGAHNVDSNEISIFIRAGEBALQKALAN 641
DB 535 EYIIPAVEKGLDSMNNGVLAGYPLVDITAKLYDSYHVDVSETKFRVAASALQKAAKN 594
QY 642 ATLCLLEPIMAVEVAVNPFQGVAGINRRRGVITGDDGVEDYFTLYADVPLNDMPGYS 701
DB 595 ANPVLLEPMKVTIIVPDDYLDGIDGHTSRRGRVGEHAGNSQIVANMVLAEVFGYA 654

RESULT 11
US-09-815-242-10481
Sequence 10481, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27

```

? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 10481
? LENGTH: 693
? TYPE: prt
? ORGANISM: Enterococcus faecalis
US-09-815-242-10481

```

Query Match	34.7%;	Score 1352.5;	DB 10;	Length 693;
Best Local Similarity	42.1%;	Pred. No. 312e-104;		
Matches 292;	Conservative 13;	Mismatches 256;	Indels 15;	Gaps 9

Qy	4	EKIRNGISAHIDSGKTTLLTERLYVTGSAIOMHEVKNKDGAMDSMELEBORGTTIQ	103
Dd	8	EKTRNIGIMAHVAGKTTTTERLYTGTGKHKIGET--HEG-ASOMDMEDOEORGITTT	64
Qy	104	SAATFTMKDKNINIIDTPGHVDTTIEVERALRVLDGAVLYLCAVGVOCTMTVNRQK	163
Dd	65	SAATTAQMKCYRVNIIDTPGHVDTTIEVORSRLVDGAVTLDOSQVEPQETVNRQAT	124
Qy	164	RNVVPFLTFTNKIDRMGNSPAPALQOMRSEKILMHTAPAOIPMGLEGNKGIIVDIIEBAT	223
Dd	125	EYKPRVLFVFNKNDKIGADFPYVESLEHRLDANAHPIQIPGAEBEDTGTIIDIKKAE	184
Qy	224	YFDODFSQIYRVEIIPAEILAAATHRQELICVANSDEOLGEMPLEEKIPISIDLKAI	283
Dd	185	IYNDLDTGDIQETDIDEDYLEKNGEHRKLVAVAFETDOLMKKYLBEGETESELVAGI	244
Qy	284	RRAITKRSFTRPVLSGALKKKGYQPLDVALEYLNPBSVQNVAILANKDSDKCKTILM	343
Dd	245	RQATINVEFFPVLAGSAFNKKGQVLDAVLDYLPSPDLIDAIKIGIDKTD--EETTR---	300
Qy	344	NSRSHNSHPVGLAFPLEVGRF--GOLTVYRSYOGELKGGDITVYTRTKRYLRLRMH	402
Dd	301	--PADDAFPASIAFLKMTMDPFQGRITLTPRVYSVLBEGSYVLANSKCKERIGRIOMH	358
Qy	403	ADMMKASTEVEVADICALFGI--DCASGDTFTDKANSGLSMESIHVPDPVISIAMKDSNK	461
Dd	359	ANTRQ--EIDKYVSGDILAAVGLKDTTGTGDTLC--ALDAPVILSESTIEFPBPVQVAVEKSK	416
Qy	462	NDLEKFSKGIGRTREDDPKFYVFDENKETYISGMEGLHLEYIAQRLEREYGPCTIGK	521
Dd	417	ADODKQVLOKLAEEDPSFRVETNVESEIYISGMELHLDVLDVNRKKEFKYEAANVGA	476
Qy	522	PKVAFRETTIAPVPDFDTHKKQSGGAGQYKYGIVLEPLDEPDYTKLEFSDETGSI PK	581
Dd	477	POVSYRETFPAATKAKGKFPVROGGKGQYGHVWEFTP--NBEVGFEEFEMAIIGVAVPR	534
Qy	582	QPVAPVKEGGLDACCEKGPLSGHGLSLRVLVODGAHHVWDSNEISFTIAGSGALKOLAN	641
Dd	535	EYIPAEVKEGLEDMMNNGVLAGPYLDIKKLVDOSYHDVDSNETAFRVAAAMALKAAAKN	594
Qy	642	ATLICIEPIAAVAVVAPNEFOGAVIAGIRRHGCVITGODGVEDVFTYLAADPLMDMGYS	701
Dd	595	ANPILIEPMMKVTITVPEDYLDIMGHVTSRRGRVREGMEAHGNQIVANVAPLAEHGYA	654
Qy	702	TELRSCBCEGSEYTMESKRYOCLPSTQDQDVNKK	735
Dd	655	TTLRSAFOGRGTMMVFDHYEDVPKSVODEIILKK	688

RESULT 12
US-09-815-242-13233

```

; CHANGED INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.

```

```

APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA, 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIORITY APPLICATION NUMBER: 60/191,078
PRIORITY FILING DATE: 2000-03-21
PRIORITY APPLICATION NUMBER: 60/206,848
PRIORITY FILING DATE: 2000-05-23
PRIORITY APPLICATION NUMBER: 60/207,727
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: 60/242,578
PRIORITY FILING DATE: 2000-10-23
PRIORITY APPLICATION NUMBER: 60/253,625
PRIORITY FILING DATE: 2000-11-27
PRIORITY APPLICATION NUMBER: 60/257,931
PRIORITY FILING DATE: 2000-12-22
PRIORITY APPLICATION NUMBER: 60/269,308
PRIORITY FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 13233
LENGTH: 693
TYPE: PRF
ORGANISM: Streptococcus pneumoniae
US-09-815-242-13233

```

Query Match	34.7%;	Score 1352.5;	DB 10;	Length 693;
Best Local Similarity	41.9%;	Pred. No. 3.2e-104;		
Matches 291;	Conservative 140;	Mismatches 248;	Indels 15;	Gaps 9

Oy	44	EKIRNIGISAHNIDSKDTTLTERVLYYGRILAKMEYKQGVAVYMDMSLEBORGITTO	103
Dd	8	EKIRNIGIAHAHDACKTTTERILYYGKIHKIEET--HSG-ASQMDMMQEOBERGTTT	64
Oy	104	SAATFTMMKDVANINTIDTPGHVDFTEVERALRYLDAVAVLCAVGSVCQCTMTVNRQK	163
Dd	65	SAATTQMMNNHVNIIIDTPGHVDFTEVERALRYLDAVAVLDSQSVETETVWRQAT	124
Oy	164	RYNVPELTETINKLDRMGSNPAPALQOMRSKLNHTAEMQIPMGIEGNFKGIVDLEERAT	223
Dd	125	EYGVPRIVFANMKDKIGADFLYSVSTLHDRLQANAHPIQPIGSEDDPFRGIIDLIKAKA	184
Oy	224	YFPGDSSQIVRGEIPAEIRPAAADTHRELLIECVANDDEQGEHFELEKPISTISDLKLA	283
Dd	185	IYNDIDGTDLBEDIPAEYLDQAOEYREBKLIBAALTEDEELIMKYLEGEBITHEELKAGI	244
Oy	284	RRATLRSEFTPELISALKNKGVOPLDPAVLEYLPNPSEVONVALNKKQDSKEKTKILM	343
Dd	245	RKATINVEFPFLPCSAFRKMGVQMLDAVIDIYPSPLDIPALGINPDTDAE----IR	300
Oy	344	NSSRHNHPVGLAEFPLEVGRF-GQLYVRSYOGELKKDGTIYNTTRTKVRLQRLARNH	402
Dd	301	PAS--DEEPFALAKRIMTDPFVGRLEFRFYSGVLDSGSVLTNSKXKERIGRILONH	358
Oy	403	ADMMEAESTEVEYAGDICALFGI-DCASGDPTTDCANGLSMESHVDPDIYSTAMKPSNK	461
Dd	359	ANSRQ-EIDTVVSGGIAAVALGKIDTTGDSLTDE-KAKIILIESINVEPIQIOMVEPERSK	416
Oy	462	NDEKRSKIGKFTDEDPFKYKYPDTENKEVVISGEMELHLEIYAQRLBEEVQCPICITGK	521
Dd	417	ADDDKKGIALQCLAEEDPFRVETNVEGEFVISGEMELHLDVLDVDMRREPFVEANVGA	476
Oy	522	PKAARETITTAVPDPDTHKKMGSGAGYGVIVLPELDPEDYTKLEFSDFTGSGNIIPK	581
Dd	477	PQVSYHETRAQTQARGCFRQSGSGKQGFV--WIEFTPNBEKGEFENALVGVVPR	534
Oy	582	QVPAVEKGFILDACEKGPLSGHLSGLFVLQDGAHHVDSNETISFIRAGEGALKQALAN	641

Db 535 BPIPAEKLIVISMANGVAGYPMVDVAKALYDGSYHDVTSSETAKFIASLSKEAAS 594
Qy 642 ATLCILEPIMAEVVAPNEFOGVINGIRRHGVITGODGVEDYFTLYADVPLNDMFGS 701
Db 595 APPALIEPMMLVTTITPEENLADVMGHVTRKGRVGMPEHAGNSQIVRAYVPLAEMFGYA 654
Qy 702 TELRSGTEKGGEYMEYSRYQPCLPSTOEDVINK 735
Db 655 TVLRASAGRGTFMVFHDYEDVPKSVQEEIICK 688
RESULT 13
US-09-815-242-12443
Sequence 12443, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12443
LENGTH: 715
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12443
Query Match 34.7%; Score 1351.5; DB 10; Length 715;
Best Local Similarity 40.3%; Pired. No. 4.1e-104;
Matches 295; Conservative 133; Mismatches 269; Indels 35; Gaps 10;
Qy 16 RAPASIGQKQVNMWACWSSSGVLPNKIRNIGSAHIDSKTTLTERVLYYGRIRAK 75
Db 3 RASPISTYSPNSINMKEKTYMAR-EPSLKEKTRNIGMAHIDGKTTTTRILYYGRIRIK 61
Qy 76 MHEVKGKLVGAVMSMELERQGITIGSAATFTMKADVINIIDTPGHVDFTEVERAL 135
Db 62 IGBT--HEG-ASQMDMBEODRGITITSAATTAAMEGRNVNIIDTPGHVDFTEVEREL 118
Qy 136 RVLDAVAVLCAVGIVQCCOTMTYNNRMKYNVPLTFIKLDRMGSNPARALQOMSKUN 195
Db 119 RVLDAVAVTLDAQSGVEPQETETWRQATTVGVPRIVFVNMKDLGANSFVSSTLDRIQ 178
Qy 196 HNTAFWQIDPMGLGKNGKIVLIEERATYFDGFSQIVYGEIIPARLRAATDHROELLE 255
Db 179 ANAAPQIDPIGADEDEBAIIDLVEEMKCFKTYNDLGEIEIEIEIFPEHDLRAEARSLE 238
Qy 256 CVANSDEQJGEMFLEBKIPISIDLKLAIRATLKRSFTVFLGSAKKNQVPLDAVLE 315

Db 229 AVALTSDEIMKCYLDEEBISSELKEAIRQATTNVEFYVPLCGTAFKXGVQJMLDAVID 298
Qy 316 YLPNSEVO-----NVALINKDDSEKKTILNNSRHNSHFVGLAPLEVGNF 365
Db 299 YLPSELDVPIIIGHRASPEEBEVINKADDSAE-----PALAFKWTDPY 343
Qy 366 -GOLTVVNSYQELKKGDITVYTRTKVRLQRLARMHADMEASTEEVYAGDICALGCI 424
Db 344 VGLTFEPVYSGTWSGVVKNSTGKRERGRLLQWANSRQ-BIDTVYSGDIAAAGL 402
Qy 425 -DCASGDTFTDKANGLSNESIHVPDIYISLAKNSKNDLEKSKGIGRFPREDPTFKV 483
Db 403 KDTGTGDTLCEKND-IILSEMEPEPVIHLSEBKSADODKMTQALVKQLEBPTFHA 461
Qy 484 YFDENKKEIVISGMELHLEIYAOGLREYGCPTGKPKAFRETIAPVPDPDTHKKQ 543
Db 462 HTBESTGOVITIGMELHLDLVDMKKEFNVECNVAGAPMVSARETTPKSSAOVGKFSRQ 521
Qy 544 SGAGQYGVIGVLEPLDPEDYTKLEFSDETFGSNIPOQVPAVAKGFLDACCKPLSGH 603
Db 522 SSGRGQYGDV--HIEFTNETGAGFEFENALVGVVPREYIPSVAGLKDAMENGVLAGY 579
Qy 604 KLSGLRFPVLDGAAHMDVSNESISPRAGEGALKALANATLCILEPIMAEVVAPNEFOG 663
Db 580 PLIDVAKALYDGSYHDVDSSEMAFIIAASLKEAKKCDPIILPIMKVTIEMEPEYWG 639
Qy 664 QVIASINRHGVITGQDVEDYFTLYADVPLNDMGYTELRSCTEGGEYMEYSRYQP 723
Db 640 DIMGVTNRGRGVDDMEPRGNQVNAVYPLSEMGVATSLRSNTQSGTITMYFDHYAB 699
Qy 724 CLPSTOEDVINK 735
Db 700 VPKSIADIEDIICK 711
RESULT 14
US-09-815-242-12736
Sequence 12736, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 12736
LENGTH: 715
TYPE: PRT

ORGANISM: Staphylococcus aureus
US-09-815-242-12736

Query Match 34.7%; Score 1351.5; DB 10; Length 715;
Best Local Similarity 40.3%; Pred. No. 4,1e-104;
Matches 295; Conservative 133; Mismatches 269; Indels 35; Gaps 10;

QY 16 RAPASIGMQRKQVWACRWSGSGVLPNEKIRNIGISAHIDSQKTLTERLVYTGRIAK 75
DB 3 RASPIVSYLFSNNWMEKKYMAR-EFSLEKTRNIGIMAHIDAKTTTERILVYTGRIAK 61
QY 76 MHEVKGKGVGAVMSDELERQGITTSQAATFMKDVNINIIDTPGHVDTIEVERAL 135
DB 62 IGRT--HEG-ASQMDMMEQODRGITTSQAATFMKDVNINIIDTPGHVDTIEVERAL 118
QY 136 RVLDGAVLVICAGVQCCMTNRMKRVNVEPLFINKLDMGNSPAPALQOMSKUN 195
DB 119 RVLDGAVLVICAGVQCCMTNRMKRVNVEPLFINKLDMGNSPAPALQOMSKUN 178
QY 196 HNTAFMOIPMGLENGFKGIVDLIEERAIYFDGDFSOIVRGEIPAEIRAAATDROBLE 255
DB 179 ANAAPLOIPGADDEFPAITDVEKMCFTKTNLGEIEIEIEIPEDHLDRAEARSLSIE 238
QY 256 CVANSDEQLGEMPLEEKIPISIDUKLAIRRAATLKRSFTFVLSALKNKGVPPLDAVLE 315
DB 239 AVAETDELMKEXLGEIEISVSELKEAIRQATTNVEFFVPLCGTAFKNKGVLMLDAVID 298
QY 316 YLPNPEVQ-----NYAILNKDQSKETKILMNSRRNSHPVGLAFLPEVGRF 365
DB 299 YLPSPIDVVKPIIGHRASNPREEVIAKDDSAE-----FAALAKVMTDPY 343
QY 366 -GOLTVYRSYOGELKGGDTTYNTRTKVYLQRLAMHADMEASTEVEYAGDICALFGI 424
DB 344 VGLTTFPRVYSGMTSGSYKNSGTGKREKREVLQMHANSRQ-EIDTVYSGDIAAAGL 402
QY 425 -DCASGDTFTDKANSGLSMESIHVPDPVISIANKPSNKDLEKFSKIGRFTREDPTFKY 483
DB 403 KDTGTGDTLCEGRND-IILSMFEPPEVILSVPEKSKADQDQMTQALVLOEDDEPTFHA 461
QY 484 YPTEKKEIVYISGMGELHLEIYVQRLEREYGCRCITGKPKVAFREITTAVPDPFTHKKO 543
DB 462 HDEEYGOVYIIGMGEHLHDLIVDRMKKEFVNECVGAPVSYREFKSSAQQKPSHQ 521
QY 544 SGAGQYGVKIVGLEPLDEPDYTKLEFSDPTFGSNI PKQFVPAVEKGLDACEKGPLSGH 603
DB 522 SGRGQYGV--HIEFTPNETGAGFEFENAI VEGVPREYIPSEVAGLMDAMENGLAGY 579
QY 604 KLSGLRFLVODGAHMYDSNEISFIRAGBGLKQALANATLCLLEPTMAVEVVAENPEFG 663
DB 580 PLIDVAKLYDGSYHVDSEMAFKTAAALAEAAKCDPVILBPMKVTIEMPEEYNG 639
QY 664 QVLAGINRRHGVITGQDGVEDFTLYADVPLNDMPGSTELBSCTGKGEYMEYSRQ 723
DB 640 DINGVTSRRGRVDGMEPRGNAOVNAYVPLSEMFGVATSLRSNTGRTTYMYPHYAE 699
QY 724 CLPSTQEDVINK 735
DB 700 VPKSIAMEDIKK 711

RESULT 15
US-09-815-242-5238
Sequence 5238, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haebelbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011a
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5238
LENGTH: 693
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-5238

Query Match 34.3%; Score 1338; DB 10; Length 693;
Best Local Similarity 41.1%; Pred. No. 5,2e-103;
Matches 289; Conservative 128; Mismatches 253; Indels 34; Gaps 9;

QY 44 EKIRNIGISAHIDSQKTLTERLVYTGRIAMHEVKGKGVGAVMSDELERQGITTO 103
DB 8 EKTRNIGIMAHIDAKTTTERILVYTGRIHIGET--HEG-ASQMDMMEQODRGITTT 64
QY 104 SAATFMKDVNINIIDTPGHVDTIEVERALRVLDGAVLVICAGVQCCMTNRMQK 163
DB 65 SAATFAAMEGHVNIIDTPGHVDTIEVERSLRVLDGAVTVDAOSGVEPQETVWRQAT 124
QY 164 RNVNPEPLFINKLDMGNSPAPALQOMSKUNHNTAFMOIPMGLENGFKGIVDLIEERAI 223
DB 125 TVGVPRIVFVNKMDLGNFEYSVSTLDRLOANAAPLOIPGADDEFPAITDVEKMC 184
QY 224 YFDGDFSOIVRGEIPAEIRAAATDROBLECVANSDEQLGEMPLEEKIPISIDUKLAI 283
DB 185 KTNLGEIEIEIEIPEDHLDRAEARSLSIEAVASSELMEKXLGEBEISVSELKEAI 244
QY 284 RPATLKRSFTFVLSALKNKGVPPLDAVLEYPNPEVQ-----NYAILNKD 333
DB 245 RQATTNVEFFVPLCGTAFKNKGVLMLDAVIDLPLSVKPIIGHRASNPREEVIAKAD 304
QY 334 DSEKTKILMNSRRNSHPVGLAFLPEVGRF-GOLTVYRSYOGELKGGDTTYNTRTKK 392
DB 305 DSAE-----FAALAKVMTDPYVGLTTFPRVYSGMTSGSYKNSITGKR 349
QY 393 VRLQRLAMHADMEASTEVEYAGDICALFGI-DCASGDTFTDKANSGLSMESIHVPDPV 451
DB 350 ERVGRLLQMHANSRQ-EIDTVYSGDIAAAGLKDGTGDTLCEGRND-IILSMFEPPEV 407
QY 452 ISIANKPSNKDLEKFSKIGRFTREDPTFKYYPPTENKKEIVYISGMGELHLEIYVQRLER 511
DB 408 IHLSEVPSKADQDQMTQALVLOEDDEPTFHAHTDEETGOVYIIGMGEHLHDLIVDRMK 467
QY 512 EYGCPCITGKPKVAFREITTAVPDPFTHKKQSGAGQYGVKIVGLEPLDEPDYTKLES 571
DB 468 EFNVECVNAGAPVSYREFKSSAQQKPSRSGRQYGV--HIEFTPNETGAGFE 525
QY 572 DETFGSNI PKQFVPAVEKGLDACEKGPLSGHLSGLRFLVODGAHMYDSNEISFIRAG 631
DB 526 NAIYGVVPREYIPSEVAGLMDAMENGLAGYPLIDVAKLYDGSYHVDSEMAFKTAA 585
QY 632 EGALKQALANATLCLLEPTMAVEVVAENPEFGQVYAGINRRHGVITGQDGVEDFTLYAD 691

Db	586	SLALKEAAKCDPVILIEPMKVTIEMPEEYMGDIMGDVTSRRGRVDGMEPRGNAQVNAV	645
Qy	692	VLNDMFGYSTELRSCTEGSGEYTWBSRYOPCLPSTOEVIK	735
Db	646	VPLSEMFGYATSLRSNTQGRGTYTWYFDHYAEVPSIAEELIK	689

Search completed: July 14, 2003, 18:27:01
Job time : 21.6298 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:11:27 ; Search time 23.0933 Seconds
(without alignments)
2620.804 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730

Sequence: 1 FCLQCTRWLRENGHFPST.....MIVKKEVSTRSASSQSSSR 2057

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:*

1: /cgn2_6/prodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/prodata/1/1aa/ECTUS.COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1492	13.9	2548	4	US-09-172-422-1
2	1274	11.9	1939	4	US-09-310-187A-1
3	1232.5	11.5	1120	4	US-09-147-404-1
4	1065	9.9	1886	4	US-08-938-105-3
5	521	4.9	103	4	US-08-905-223-395
6	265.5	2.5	3248	1	US-08-353-700-1
7	265.5	2.5	3248	5	PCT-US95-16216-1
8	264	2.5	2482	1	US-08-328-254-6
9	255	2.4	1507	3	US-08-929-329-5
10	253.5	2.4	1162	4	US-08-728-323A-2
11	253.5	2.4	1162	4	US-09-298-568-2
12	244.5	2.3	1898	1	US-08-056-200-94
13	244.5	2.3	1898	2	US-08-800-644-94
14	218	2.0	717	4	US-08-910-925-1
15	211.5	2.0	1375	4	US-09-722-139-2
16	211.5	2.0	1375	4	US-09-721-832-2
17	211.5	2.0	1375	4	US-09-721-689-2
18	210.5	2.0	1360	4	US-09-393-569-2
19	208.5	1.9	2154	2	US-08-841-349-4
20	207	1.9	1279	4	US-09-724-517-2
21	207	1.9	1279	4	US-09-641-807A-2
22	206.5	1.9	1279	4	US-09-723-096-2
23	206.5	1.9	1200	3	US-08-840-006-5
24	205.5	1.9	10182	4	US-09-134-001C-3159
25	205.5	1.9	743	4	US-08-910-925-3
26	199.5	1.9	740	1	US-08-357-073-5
27	197.5	1.8	514	2	US-08-960-022-14

28	197	1.8	1354	3	US-08-685-871-2	Sequence 2, App11
29	195.5	1.8	703	4	US-08-910-925-4	Sequence 4, App11
30	192	1.8	619	1	US-08-465-746-2	Sequence 2, App11
31	192	1.8	619	1	US-08-214-164-2	Sequence 2, App11
32	192	1.8	619	2	US-08-467-852A-3	Sequence 3, App11
33	192	1.8	619	2	US-08-246-636-2	Sequence 2, App11
34	192	1.8	619	2	US-08-247-491A-3	Sequence 3, App11
35	192	1.8	619	2	US-08-319-795-2	Sequence 2, App11
36	192	1.8	619	2	US-08-468-985-2	Sequence 2, App11
37	192	1.8	619	3	US-08-312-949-2	Sequence 2, App11
38	192	1.8	648	1	US-08-072-070-2	Sequence 2, App11
39	192	1.8	648	1	US-08-469-434-2	Sequence 2, App11
40	192	1.8	648	1	US-08-214-222-2	Sequence 2, App11
41	192	1.8	648	2	US-08-467-852A-2	Sequence 2, App11
42	192	1.8	648	2	US-08-468-718-2	Sequence 2, App11
43	192	1.8	648	3	US-08-247-491A-2	Sequence 2, App11
44	192	1.8	648	3	US-08-446-201-3	Sequence 3, App11
45	192	1.8	695	1	US-08-127-499A-23	Sequence 23, App11

ALIGNMENTS

```
RESULT 1
US-09-172-422-1
; Sequence 1, Application US/09172422A
; Patent No. 6300485
; GENERAL INFORMATION:
; APPLICANT: Adams, Arwen E.
; APPLICANT: Chiu, Choi Ying
; APPLICANT: Dahl, David
; APPLICANT: Gorman, Susan W.
; APPLICANT: Leng, Song
; APPLICANT: Sheffield, Val
; APPLICANT: Welch, Juliet
; TITLE OF INVENTION: MYOSIN IXA AND CYCLIC NUCLEOTIDE GATED
; TITLE OF INVENTION: CHANNEL-15 (CNGC-15) POLYNUCLEOTIDES, POLYPEPTIDES,
; FILE REFERENCE: 200130.442
; CURRENT APPLICATION NUMBER: US/09/172,422A
; CURRENT FILING DATE: 1998-10-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 2548
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-172-422-1
Query Match
Best Local Similarity 24.3%; Pred. No. 5.2e-107;
Matches 541; Conservative 349; Mismatches 674; Indels 660; Gaps 79;
13.9%; Score 1492; DB 4; Length 2548;
35 YGVFTYKQSTTHQKVA--MPTNEGVDMASLTLEHGSIMVNLFOYKKNQIWT 91
118 YGSLGSLVLRVTERRRMRGRFLPQPOKDFDLCSLPINKETLENLDRFKHEKIYT 177
92 YIGSLIASNPQPLAGYEPATMEQYSRRHGLPPHIFALANECYRLMGRHNOCLT 151
178 YGSLILYINPK-FLPIYKPYVMYDNDHKGKEPHIYAVADVAYHAMLQRKKNQICV 236
152 IKGSGAGKTEBTKILKPLSVISQOSLELSLKETSCEVRAIILESSPIMEAFNAKTYV 211
237 IKGSGSGKGTGNTNLHLHILTLSSQGF-----ASGVEQIILAGPVLERFGAKTNH 269
212 NNNSRFGKPVQNLTCQKNIQGRIVDCILSSQNRVVRQNGENNYHIFVYLLAGLEHE 271
290 NNNSRFGKPIQVNOQETVLGA-YVERKYLLEKSRILVQEHENRHHYHVFYLLAGASD 348
272 EEEERYLSPREYHNLN-----SGVEEK-----TISDSE-----SPREVTAMDV 313
349 EKSAHKLQPEEYHYNLQITKKPLRQSWDDYCYDSEPCFVLEGDLRHDFRLQALMEW 408
```


QY 314 M0P8KEVEVSHLAGILHAGNI-----EFTAGAOVSFKTALGRSABLLGLDPTQILD 369
 Db 409 VGLPRTBROIFSLLEAILHAGNI CYKKKTYRDSIDICNPEVLPIVSELLEVEKEMLE 468
 QY 370 ALTORSMPLRGEILPLPLAVQAVDSRDSIALMALLYACCEBWTIKTINSRKINEDK--- 426
 Db 469 ALVTRTYVAGEKLIIPYKLAELAVTRNSMAKSLYALPDWIVFRINHLLMSKDEHNT 528
 QY 427 ---SIGILDFGENEFVNHFEQFNINYANEKLOEFKHKHIFSLDEULEXREGLVEDID 483
 Db 529 KTLISGLVDLFGEDIEENNSFEQFCINPANELOHFNQHIFYLEQEBETREGISHAND 588
 QY 484 WIDNGECLDIEKK-LGLLALINESHPQATDSTLEKLSQHANNHFVYKPRVAVNPF 542
 Db 589 YIDNTCCINLISKKPTGLHLHDESNFPOATNQTLIDKFKHQHEDNSYIEFPVAVNEPAF 648
 QY 543 GVNHVAGEVOYDVRGLERKRDTPRDDLLNLRESFDFYDL----- 585
 Db 649 IKHVGAKVYGVYDREKNTDHRPDIVALLRSSNAFISGMIGIDPVAVFRMAILRAV 708
 QY 586 -----FEHVSNN-----NODTLKCG----- 601
 Db 709 FRAMAFAFRAGKKNIRKGTGHDTAPCALIKMSDFSLOHPVQROSLBILQRCKEKYS 768
 QY 602 ----- 601
 Db 769 ITRKNPRTPLSDQNMALNEKNOHDTPIAMNGRTGIRQSRLSSGTSLLDKGIFANST 828
 QY 602 -----SKRRAP 607
 Db 829 SSKLTERAHGILTRNNKPKSKPALPHLEEVNSLKLTRLTLQDRITKSLHLHKKKKRP 888
 QY 608 TVSSQKVDLSHLMATLSSSNPFYRCIKPNQKMPDQDAVAVNOLRYSGLMETVRI 667
 Db 889 SISAOQA-SLSTLMTLGOAEFYFKCIRSNABKPLKRSVDVLRQLRTSMLETVQI 947
 QY 668 KAKYAVRRPPODYRRKYVLMENLALPEDVRGKCTSLQLYDASNEWQLOKTKVFLRE 727
 Db 948 ROGVSYSKSFODFVSHFVFLPRNIIPS--KENIQDPFKINLNDPNYQVGTWVFLKE 1005
 QY 728 SLEQLEKREBEV-----SHAANVTYRAHVTGFLARKOYRK- 763
 Db 1006 QERQHLQDLHQLVLRRIILQRMFVULLCROHFLHLRQASVLIQFMFNRYLNQKQVRDA 1065
 QY 764 -----VLYCVIIOKRYAFILRRRFLHKKKAIIVFOKORGLIAR----- 805
 Db 1066 AVOKDAFWASAAALLOASWRAHLERORYILBPAALIVIOKKRDYRRRRHMAITCIQAR 1125
 QY 806 --VYRQLLAEKREQEBKKQOEBEKKKREBEREREREREAELRAQOEB-----ETRK 857
 Db 1126 WKAYRE--SKRYQOQRKIIILQSTCRGFARQFALKQOERETKPEVGLVNIKGYG 1182
 QY 858 QOELHLOKSO-KEALITRELEKOKENKQVEILREKEIEDQNKKE-----QOELSLRE 912
 Db 1183 SLEIQSDPSEWEDCSFDRIRIKALIECKSVITESNRISRE--SSVDCIKESPNKQOEAQSOQ 1241
 QY 913 ASLOKLOE-----RQDELRLREEBACRAQOEFLESINPDEIDECYRNI-----RSLSG 963
 Db 1242 SGVD-LOEDVLRERERSLEDLHQKVGAKR--ESRRKRELQALFSLBLKVRGLGI 1298
 QY 964 S-----EFSSELAESACER--KPNFNFSQ-----PYPEEB-----VDEGEFADD 1001
 Db 1299 SPSEDRMTSTELVPEGLQSPRGTPDESQSLLELTSYESQSKLESYISDBG-----D 1353
 QY 1002 AKRDSNPBEHGSDOPTSGIRSDSSF-----EDYVMDTVYTPSPASDSTVLAPSVQ 1057
 Db 1354 LOPSPSKISSPFRDNRALASNSNTSSAHLKODTMEKVVCSSES-----ITCKPQOK 1409
 QY 1058 DSGSLNSSSGESTYCMQONAGLPPDGDYVDODDYDGAATSGSVTFSSVYSQMS 1117
 Db 1410 DS---FISNLPFFFIPOODPLKTMISQDLTISQKRLLENBTJAEALITLDINRETR-- 1464
 QY 1118 PDYRCS-----VGTYNSSGAYRPSSEGAOSSPDESEDEPDSRPTDDEL----- 1161

Db 1465 -RYHCSGKQOIVPSINTESSNPVLKKEKLENTKEERQKOLQOONKEKEMEOIROOTDIL 1523
 QY 1162 -----SYRDSYVSCVTLPYFHSFLYMKGGLNMSWRKRCVLDFTFLFRKOE-- 1211
 Db 1524 EKERRAKFTIEPRIGECIVAP-----SSYOSKORVERPSSLSLNTSNKE 1570
 QY 1212 -----ALKQGLMKKGGSGSTLSRRNWKARFVLRQSKLMYFENDSEBKKTIV----- 1260
 Db 1571 LNVLSGLSKDAALAQOSSSAHLPPKD-----RVTYFFERKSPQOSSYKELS 1621
 QY 1261 -----EVRNAKEIIDN-----TTKEN-GIDI 1280
 Db 1622 KTDNRGTOQLNVAACKLSNNRISKREHFRPTQSYSHNSDDLREGNARPIFFTQKDMNSIPL 1681
 QY 1281 IMADRTFHLIASP-----EDASQW--FSLQVHASTQEOIEQMHDEQONPQNAVGLDV 1334
 Db 1682 VKKA--LNSKNPOLKHEDEPAMKPVKLAGEQETQGRFSSV--DEQAKLHK--TWSQ 1734
 QY 1335 GLIDSVK-----ASDPDRPNSEFVITANRVLHCNADTEEMHWTLLQRSKGTRE-- 1388
 Db 1735 GEITLKAARQKASDSDIRQR-----AKRFPKAGQGEKKTIRVKT 1777
 QY 1389 -----GOEFT-----VNGWLHKEYVNSPKMSLKLKRW 1417
 Db 1778 TQSEVSPLEAGTDVIPAQFPDELAAYHTPPLSPBEPSCSCKEPEKNEKESP 1830
 QY 1418 FVLTHNSLDYVSSSKMLKLGTLVNLSCSVVPPDEKIFKSTGYVNTVYRKHCYRLX 1477
 Db 1831 -----KAKRRSVKISNVALDSM-----HMQ----- 1851
 QY 1478 TKLNEATSWSSVIONVDTKAPIDTPQOLIQDI-KENCILNSDVEQIYKSN-PILRYT 1535
 Db 1852 -----NDSVQ---IASVBDLKS-MDEFLKKYNDLNDSDSKDQDLVDVYFKKALKEFRN 1903
 QY 1536 HPHLSPPLPLPYGINLNLKQKGYTLQDEAIKIFNSLOQLESMSD-PIP1IGILOT 1594
 Db 1904 IFSFYSSALAMDG---KSIRYKDLVALFEQILEKTRMLEORDSLGESPVAV-----W 1953
 QY 1595 GHDLRPLDELYCOLIKON-----IVP-----HPSVGNLYSQIILT- 1632
 Db 1954 VNTFVFLPDE-VNNEFKTSDCTATVPKTERKRRKKETDLVEBHNGHIFKATQYISIPY 2012
 QY 1633 CLSCTFLPSRGILKYLKRLKRIEQPCTEMEKALFTYESLKT-KC-----R 1681
 Db 2013 CEYCS-----SLWIMDRASVCKLCKYACHKKCKLKTATACSKKYDELISR 2059
 QY 1682 EF-----VPSRDE-----IEALHROEM--ISTVCHGGSCCKITINSHTAGEVE 1726
 Db 2060 QGVELSLRSTSEDRTPVPLVEKLIIVYIMHGLYTEGIVKSGSTNKI-----K 2107
 QY 1727 KLIRGLAME-DSRNNFALFEVNGHVDKALIESRTVADVLAKREKLAASEVGDLPWKFY- 1784
 Db 2108 ELRQGLDTPAESVN--LDDYNIH-----VIASVFKM-----LKDLEPNPLMT 2147
 QY 1785 FPLY 1788
 Db 2148 FELY 2151

RESULT 2
 US-09-310-187A-1
 ; Sequence 1, Application US/09310187A
 ; Patent No. 6358751
 ; GENERAL INFORMATION:
 ; APPLICANT: Benichou, Gilles
 ; APPLICANT: Fedoseyeva, Eugenia
 ; TITLE OF INVENTION: Involvement of Autoantigens in Cardiac
 ; TITLE OF INVENTION: Graft Rejection
 ; FILE REFERENCE: UCSF-090
 ; CURRENT APPLICATION NUMBER: US/09/310,187A
 ; CURRENT FILING DATE: 1999-05-12
 ; NUMBER OF SEQ ID NOS: 3

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1939
; TYPE: PR
; ORGANISM: Homo sapiens
US-09-310-187A-1

```

```

Query Match      11.9%; Score 1274; DB 4; Length 1939;
Best Local Similarity 24.8%; Pred. No. 3.7e-90;
Matches 452; Conservative 334; Mismatches 640; Indels 396; Gaps 59;

```

```

27 GIVRTDYGVGYTYKOSTITTHOKTAMPTNEBEGDDNLSLTELHGSIMNLTQRYK 86
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
57 GVIATEBNGKTVTVVEDQLQO---NPKFDKQDMMLTFLHEPAPVLNFKERYVA 111
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
87 NOIMTYGSLASVNPYOPAGLYEPATMEQYSRHLGSLPHIPALAECTRCMLKRD 146
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
112 WMITYYSGLFCVTVNDEYKLP-VYNAEVAAYRGRKRSEPPHIFISDNAYQYMLTDR 170
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
147 NCCILIKGSGAGKTESTKLILKF--LSVISQOSLELSIKETSCVEPALESSEPI 203
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
171 NOSILITGSSGAKTNTKRVIOYFASIALIGDRCKDNANANKGLEQIIOANPAL 230
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 FENAKTVNNSSRFCKFYQVQNIQCKNIOGGRIYDCILSQNRVVRQNPGERNTI 263
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
231 FENAKTVNRDSSRFCKFIHFGATGKLASADI-ETYLLEKSRVIFQLAERNYHIF 289
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
264 ILAAGLEHEREEFYL-STPENHYLNQSGCEDKTSIDQSPREVIYANDWQFSKE 322
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 ILSNKKPELIDMLLVTNNDYDFAVSO-GEVASASIDSEELMADSAFDVIGFTSE 348
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
323 EVSRLLAGILHGNIEF-----ITAGGAQVSFXTALGRSALLGLDPTQLTAL 374
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
349 GYKLTGALMTHGNMKFKQKREBQAPROGTEDAK-----SAYLMGLASADLL 403
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375 SMFLNGEELTPLANVOAVDSRDSLAMALYACCFEYVIRKINSRIKGNEDFK-SI 433
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
404 RYKVGREYVTKGQSVOQVYYSIGALAKAYEKFNNMTKINATLETQPROGFI 463
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
434 PEPENEVNHPEOFINIVANEKIQEYFNHGISLBOLEYSRELVWEDIDW-IDNE 492
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
464 APELFDPSFEOQLCNFTNEKIQEYFNHMFVLEBEEYKKEIETFTFGMDLOAC 523
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
493 LLEKLGILALINEESHFOATDSTLEKLSQH--ANNHFYVKRVAVN----- 545
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
524 LLEKPMGINSIEEBECMFKATDMTKALYDNHLKSN--FOKRNKKGQEAHFSI 581
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
546 HYAGEVOYDVRGILEKRDTPRDDLLNLLRESRFDIYDLFEHSSRNQDTLKC-G 604
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
582 HYAGTVDNILGLLEKRDPLNETVALYQKSLKLMATLFSGYATADTGDGSKG 641
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
605 RR---PTVSQKVDLSHLMATLSSNPFVRCIENQKMPDQDQAVLANQLARYS 661
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
642 KSSSFQYVALHR-ENLNLKMLNLTHTTHFVACIIINERKAGVWVNDLVNHLQ 700
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
662 LETVIRIKGAVYVRFPDQFYKRYKYLMLNLALPE---IVRGKTSLLQLDVAS 717
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
701 LBGIRICRGFPNRIIYGPFRQRYRL-NPVALPEQFIHSKGTCKLSSLDIDHNO 759
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
718 LCKTVFLRESLEQKLEKREBEVSHAANVIRAHVGLFRAKOYKRVY---CV 774
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
760 FGHTVFVFAAGLGLLEEMRDERLSRIITRMQOAGQMLRIEFKIVERRDAL 819
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
775 YRAFL-----LRRFLHLK 789
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
820 IAFPMQVPMWPKLYFKI KPLLSAETEKEMATKEEGRIKETLEKSEARKKELE 879
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
790 AATVPOK---OLRGQ-----IARVYRQLLAKREOEKKOE--- 826
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
880 VSLQEKNDLQLOVQAEQDNLANDEBRCOLLNKI--QLEAVYKEMERLEBEEM 937
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
827 ---EKKRREERERERREAEELR-AQOEE-----TRKOE 860
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

938 LTAKKRLEDCSEBKIDIDLETLAKVEKEKATENKVNKLTBEVAGLDEIITKE 997
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
961 LEALQSOKEALBTELEKQENKQVEBILTEKIEBLOQMEQO----- 906
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
998 KKAQEAHQALDLDQVEDKVNLSKSKVLEQOVDLESLEBKRYRMDLEBAKRL 1057
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
907 --ELSTFASLQ-----KLQERRDQELRL 929
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1058 EGDLLTQESIMDLNDLQLEEKLKKEPDIINOQSKIEDEQALALOLOKLAKE 1117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
930 EE-----BACRAQOFLESNFPD---EIDECYRNIERSLGGSEBSFSLASACE 981
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1118 EELBEELAEKRAKRYKVLSDLSLELEISERLEEA-GGATSVQIEM---NKK 1171
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
982 NFSQPYE-EVYDEGEFADDDAF--KDSNPSEHGS-----D 1016
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1172 EFQKRRLDEETLQHEATLAALRKHADSVAELEQDINDIQRVYKLEKSEKEL 1231
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1017 QRTSGI-----RTSDSSE-----DPYNDVVTSPSADSTVLLAP 1054
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1232 DVTSMNEQIIRAKANLEKVSRTLEDQANEYKLEBAQRSINDFTQRAKLQTE 1291
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1055 SVQDSGSLHNS-SGSESTYCHPN-----AGDLPSFGQDIDYDODDED 1097
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1292 QLEKEBALISQLTRKGLSYTOQMEDLKQLEBEGKAKNALAHALQSAHDCDL 1351
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1098 GATSGS---SVTPENSYGQSPDYRCVS--GTVNSGAVRFSBGASQSPDESD 1152
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1352 ETEAAEELQRYSKANSVAQRTYETDIALQRTLEBEAKKLAQSLQDA-E 1410
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1153 SRFDTDELISYR-----RDSVYSGVTLPYHPSFLYMKGLNNSYKRW--- 1195
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1411 AKCSLEKTRKRLQNEIEDLVNDVRSNAALADKKQNF---DKILAEMKQY 1466
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1196 -----CYLQKETFLMRSKQALKQCHLHK-----GGSSSTL 1228
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1467 SELESSQKARSLSLEFLKNAVESELEHETFGRENKNOBESISDLTEQ 1526
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1229 SRNWKQKMFYLR---QSKL-----MYFENDSEBKGTVEYRTAKI 1281
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1527 ELEKRYQLEVEKLEQSLAEBAEASLEHESKILRAQLEPQIAELIRKLA 1586
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1282 MADRTFHLIASEPE---DA-----SQWPSVLSQVNAS-TDOEIQME 1330
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1587 QAKRNHQVNVSLQSLDAETRSRNEVLRYKKMGEDLNEMEIQLSHANRMA 1646
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1331 TL-----DVLGIDVCAQSDSPDRPSFVITIANRVLHCNADTBEBNHN 1383
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1647 SIQSLKDTQIOLDVAHANDL-KENIAIVERNNLLQAEI---EELRAV 1701
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1364 DTRVGOEPIYRGWLHKEVNSPKMSILKTKRWFLVTHNSLDYKSSBKXAL 1443
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1702 -KLAEQELI-----ETSERVQLHSQNT-----SLNQKKNESD 1744
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1444 NSLCGVNPPDEKIFKETGYMNVTVYGRGHYRLYTKLINEATRWSSV 1503
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1745 EAVQCRNAEBA-----KCAITDAMMAEUKKEDDTSAHLER 1783
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1504 PTQQLIDIKENCNSDVVEQI 1525
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1784 MKNMEOITKDLQHLDEAEQI 1805
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

RESULT 3
US-09-147-404-1
; Sequence 1, Application US/09147404
; Patent No. 6326163
; GENERAL INFORMATION:
; APPLICANT: FORSMANN, WOLF-GEORG
; APPLICANT: RAIDA, MANFRED
; APPLICANT: BRENNER, BERNHARD
; APPLICANT: NIER, VOLKER
; TITLE OF INVENTION: A METHOD FOR THE DIRECT DIAGNOSTIC DETECTION OF

```

TITLE OF INVENTION: GENETICALLY CAUSED PATHOGENIC POINT MUTATIONS
FILE REFERENCE: 10466/P63231US0
CURRENT APPLICATION NUMBER: US/09/147,404
CURRENT FILING DATE: 1999-04-06
NUMBER OF SEQ. ID NOS: 3
SOFTWARE: PatentIn Ver. 2.1
SEQ. ID NO. 1
LENGTH: 1120
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: MOD RES
LOCATION: (26)
OTHER INFORMATION: "Xaa" represents Ala or Val
NAME/KEY: MOD RES
LOCATION: (59)
OTHER INFORMATION: "Xaa" represents Val or Ile
NAME/KEY: MOD RES
LOCATION: (143)
OTHER INFORMATION: "Xaa" represents Arg or Gln
NAME/KEY: MOD RES
LOCATION: (245)
OTHER INFORMATION: "Xaa" represents Arg or Gln
NAME/KEY: MOD RES
LOCATION: (256)
OTHER INFORMATION: "Xaa" represents Gly or Glu
NAME/KEY: MOD RES
LOCATION: (403)
OTHER INFORMATION: "Xaa" represents Arg, Gln, Leu or Trp
NAME/KEY: MOD RES
LOCATION: (453)
OTHER INFORMATION: "Xaa" represents Arg or Cys
NAME/KEY: MOD RES
LOCATION: (513)
OTHER INFORMATION: "Xaa" represents Phe or Cys
NAME/KEY: MOD RES
LOCATION: (584)
OTHER INFORMATION: "Xaa" represents Gly or Arg
NAME/KEY: MOD RES
LOCATION: (587)
OTHER INFORMATION: "Xaa" represents Asp or Val
NAME/KEY: MOD RES
LOCATION: (602)
OTHER INFORMATION: "Xaa" represents Asn or Ser
NAME/KEY: MOD RES
LOCATION: (606)
OTHER INFORMATION: "Xaa" represents Val or Met
NAME/KEY: MOD RES
LOCATION: (615)
OTHER INFORMATION: "Xaa" represents Lys or Asn
NAME/KEY: MOD RES
LOCATION: (716)
OTHER INFORMATION: "Xaa" represents Gly or Arg
NAME/KEY: MOD RES
LOCATION: (719)
OTHER INFORMATION: "Xaa" represents Arg or Trp
NAME/KEY: MOD RES
LOCATION: (723)
OTHER INFORMATION: "Xaa" represents Arg or Cys
NAME/KEY: MOD RES
LOCATION: (731)
OTHER INFORMATION: "Xaa" represents Pro or Leu
NAME/KEY: MOD RES
LOCATION: (736)
OTHER INFORMATION: "Xaa" represents Ile or Met
NAME/KEY: MOD RES
LOCATION: (741)
OTHER INFORMATION: "Xaa" represents Gly, Arg or Trp
NAME/KEY: MOD RES
LOCATION: (778)
OTHER INFORMATION: "Xaa" represents Asp or Gly
NAME/KEY: MOD RES
LOCATION: (797)

OTHER INFORMATION: "Xaa" represents Ala or Thr
NAME/KEY: MOD RES
LOCATION: (870)
OTHER INFORMATION: "Xaa" represents Arg or His
NAME/KEY: MOD RES
LOCATION: (908)
OTHER INFORMATION: "Xaa" represents Leu or Val
NAME/KEY: MOD RES
LOCATION: (924)
OTHER INFORMATION: "Xaa" represents Glu or Lys
NAME/KEY: MOD RES
LOCATION: (935)
OTHER INFORMATION: "Xaa" represents Glu or Lys
NAME/KEY: MOD RES
LOCATION: (949)
OTHER INFORMATION: "Xaa" represents Glu or Lys
US-09-147-404-1
Query Match 11.5%; Score 1232.5; DB 4; Length 1120;
Best Local Similarity 30.4%; Pred. No. 2.7e-87;
Matches 338; Conservative 216; Mismatches 374; Indels 183; Gaps 33;
8 VMLRENGOHF-PSYNSCAEGIVPRTDYGQVFTYKQSTITHQKTAHPTNEGVDDMA 66
37 VFPPDDKQEFVAKIVSRBEGKXKTAETEGKTVTKEQVWQO---NPPKFKIEDMA 91
67 SLTEHSGSIMYNLFQRYKRNQIMTYIGSILASVNPQPIAGLYEPATWEOYSRRHLGEL 126
92 MTFPLHEPVLVNLDRISQSMITYSGLFCTVAPYKLP-VYTPVVAAXKKGRBSA 150
127 PHIFAIANECCYRCIMKRDHQCILIKBEGAGKTESTYLILKPLSVISQSELSLEK 186
151 PHIFISIDNAYQVMLTRENQSIITGESGAKTWTKRVIQYPAVIAAIG-DRSKKQ 209
187 T---SCVEBALLESSPIBEAFGNATVYNNNSRRGKFPQNLICQKNIQGRYDCLIS 243
210 SPKGKLTEDQIIOANPALEAFGNATVYNNNSRRGKFPQNLICQKNIQGRYDCLIS 268
244 SONRVVRQNGERNHIFAYLAGLEHEBEREFY-STEPENTHYLNQSGVEDKTSIDQE 302
269 EKSRIYFQKABRDHIFQILSNKKPELDMLLTNNPYDAFISQ-GETTVASIDDAE 327
303 SFREVITAMDVQSKKEVEVSRLLAGILGNIEF-----ITAGAAVSFKTALG 354
328 ELMATDNMFADVGLTSEENKSMYKLTGAIMFGNKKFKLKQREQAEPDGTSEADK--- 383
355 RAABELGDPDTQALDQSGMFLRGEELITPLANOQAVDSNDSLAMALYACCFPMVIRK 414
384 -SAYLMGLNSADLLKGLCHPXKVGNEVYTKGQNVQVYATGALAKAVTERMFMMVYTR 442
415 INSRIKGNEDPK-SIGILDIFGFENFEVNHFEQFINYANEKLQEVFNKHFSLBQLEYS 473
443 INATLETQPKXYFYGVDIAGFEI.FPDNSPFQCLINFNEGLQOQFNNHNVVLQOEBEK 502
474 REGIWEEDID-WIDNGECLDLEKKLGLLALINESHPPOATSTLLEKLSQH---AN 528
503 KEGIEWTFIDKGMDOACIDILEKPMGIMSLIEECMPFKADMTFKALKFNHNGKSN 562
529 NHFYKPRPAVN---NGCVGHYAGEVOYDVGILEKRDTRDDLNLILRSRDFIYD 584
563 ---FOKPRNIGKPEAHPSLIHYAXIVXNIIIGWLQKNDPLKEXVGLGYQSSSLXLLST 619
585 LFEHVS-----RNQDTLKCGSKRRRPTVSQFVDSLSHSLMATLSSNPFVCIKPN 639
620 LFAANYAGDAPLEKGRKAKGSSFO--TVSALNR-ENLNLKMTNLBRTHPHFVACIIPN 676
640 MQKMPDQFQAVVNLQRLVSGMLFTRIRKAGYAVRPFQDFYKXYKVMRLNALPE--- 696
677 ETKSPGVMDNPLVMHQRLKNGVLREGIRICRKGFPRIILYDXQXYXIL-NPAALXEGGF 735
697 -DVRGKCTSLQLYASNSWQLGKTVFLRESLEQKLEKREEVSNAAVYRAHVLCF 755
736 XDSRKAEBKLSLSDIDHNQYKFGHTKVFPKAGLLGLLEMXEBRLSRIITRIQASRGV 795

QY 756 LARKOYKLVY---CVIIIOKNYRFL-----LRRR 783
 DB 796 LKRMEXKJLBERDSLVIQWNIIRAFMGVWMPMKLYFIKIPKLKSAREKEMAMKBE 655
 QY 784 FLHLKA-----AIVFQK---OLRQO-----IARVYR 808
 DB 856 FRLKALKKSEARKKELEBKWVSLLOEKNDLOLQVQAEQDNLADAEKCDQXIKKI-- 913
 QY 809 OLLAEKREGEKKOEK---EKKREBE-----ERERERRRA 845
 DB 914 QLEBAKEMXKRLDEEEMNAKLTAKKRLKEDCSYLKRDIDLELTAKVEKXATN 973
 QY 846 ELRAQOEER-----TRKOELALQSKQKAEKLTRELKQKQKQVEIILREKIED 898
 DB 974 KYKNIITREMAAGDEIIAKLTKEKKALQEAHQALDLOAHEDKVNLTAKYKVLBOQVND 1033
 QY 899 LQRMKEQO-----ELSTREASIQKQERHQERLRELEBACRAQOEFL 941
 DB 1034 LEGSLEQOEKVKRMDLERAKRKLKEDLKTQESIMDLENDIKQIDERLKK-----KDFEL 1087
 QY 942 ESINPDEIDECVNIERSLSGSEFSSELAE 972
 DB 1088 NALN-----ARIEBOALSGOLOKJLKE 1110

RESULT 4
 US-08-938-105-3
 Sequence 3, Application US/08938105
 Patent No. 653151
 GENERAL INFORMATION:
 APPLICANT: Leinwand, Leslie A.
 APPLICANT: Vikstrom, Karen L.
 TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
 NUMBER OF SEQUENCES: 3
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Sheridan Ross P.C.
 STREET: 1700 Lincoln St., Suite 3500
 CITY: Denver
 STATE: CO
 COUNTRY: U.S.A.
 ZIP: 80203
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/938,105
 FILING DATE:
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Crook, Mannel M.
 REGISTRATION NUMBER: 31,071
 REFERENCE/DOCKET NUMBER: 3595-4
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (303) 863-9700
 TELEFAX: (303) 863-0223
 INFORMATION FOR SEQ. ID NO. 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1886 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-938-105-3

Query Match 9.9%; Score 1065; DB 4; Length 1886;
 Best Local Similarity 22.9%; Pred. No. 7.9e-74;
 Matches 420; Conservative 321; Mismatches 622; Indels 470; Gaps 57;

QY 27 GIVVPTDVGQVTVYQSTITHQKVTAMPTNEGVDDMASLTLLGSGSIMYLPFRYK 86
 DB 56 GKVTAEITENGKTVTVKEDQVMOQ---NPPKFDKIEDNAMLTFLHEPAVLNFKERYA 110

QY 87 NOIWTVGSIILASVNPYOPIAGLYEPATMEQYSRRHLGELPHEIAJANEYCRLMKZHD 146
 DB 111 MWIITYSGIFCTVYVPIKMLP-VYNAEYVAAYRGKREBAPPHISIDMAQWLTJRE 169
 QY 147 NOCIILKBSGAGKTESTFLIKF--LSVISQSLSLKETSCEVERAILLESSPIMEA 203
 DB 170 NOSIILITESGAGKTVNTRKRVYQFASIAAGDRSKDNPANAKOTLEBOITIOANPALBA 229
 QY 204 FGNATVYNNNSRRGKVFQVNICQKNIQGRYUDCLTSSGNRVFQNGPGRNHIIPA 263
 DB 230 FGNATVRNDSRRGKFRIRHFGATGKLASADI-ETVLLKSRVIFOLKAKERNHIIFQ 288
 QY 264 ILIAGHEHEEREFYV-STPENHYINQSGVEDKTIQDSERREVIYTMADVMQFKEEVR 322
 DB 289 ILSNKKPELDMLVLTNNPYDYAFVSQ-GEVVASIDISELLALDASFDVIGFAEBEA 347
 QY 323 EVSRLLAGILHIGNIEF-----ITAGQAVSFKTAGRSABLLGHDPTQLTALTOR 374
 DB 348 GYVKLTGAIMHGNMKFKQKQREQAEPPDGEDADK-----SAYIMGINSADLLKGLCHP 402
 QY 375 SMFLAGEELTPLANQAVDSRDSLAMALYACCFEYVTKIINSRIKCNBDFK-SIGIIDI 433
 DB 403 QYKVGNEYVTYKGQSVQVYYSIGALAKSVYKEMFMWYTRINATLETYQPRQYFPGVLDI 462
 QY 434 FGFENFEVNHFEQFNINYANEKIQEYFNHIFSLBQLEKREGVMEHDWIDNGECDL 493
 DB 463 AGFE-----ISSLPHL----- 473
 QY 494 IEKRLGLALLINESHPQATDSTLEKLSQH--ANNHFFVKRVAVN--NGVGH 546
 DB 474 --KLINGISILIEECMPKATDMTFKALYDHLKSN--FOKRRIVYKQEAHFSLVH 529
 QY 547 VAGEVQVVRGTLKCRDPTFRDNLNLRSEFDPYTLDFEVSRRINODT-----LKC 600
 DB 530 YAGTVDYNIILGWLEKNKDPLETVGLVQKSLKLMATLFTSYASADYDGSKGKGGKK 589
 QY 601 GSKHRPVYSQFKVDSLHSLMATLSSNPFVRCIKPMQKMPQPOAVLANLOARYSG 660
 DB 590 GSSFG--TVSALHR-ENLTKMTNLTHTPHVRCIITNERKAPGVMDNPVLMQLRGNG 646
 QY 661 MLETVRIRKAGYAVRPFQDPYKRYKVLNRNLALPE---DYRGKTSLLQLYDASNSEW 716
 DB 647 VLEGIRICRKGPPNRIILYGDFFQRYRIL-NPAIRPEGFIDSGKAPKLGSLDIDHNY 705
 QY 717 QUGTKVYLRBSEBQLEKREBEVSHAMVIRAVLGFARKQVRILY---CVIIIOK 773
 DB 706 KFGHTKVFVKAGLGLLEBMRDERLSRIITRIQARQQLRIEFKQVVERRDALLVITQW 765
 QY 774 NYRAFL-----LRRRFLUK 788
 DB 766 NIRAFMGVWMPMKLYFIKIPKLKSAREKEMAMKBEGRVADLKSARKRELEBK 825
 QY 789 KAIIVFQK---OLRQO-----IARVYROLLAEKREKKOEK--- 826
 DB 826 MWSLLOEKNDLOLQVQAEQDNLADAEKCDQXIKKI--QLEAKYKZMTERLEDEEMNA 883
 QY 827 ---EKKREBEREREREREAEIR-AQOEER-----TRQO 859
 DB 884 ELTARKRLDECSLKDIDLELTAKVEKXATENKYKNIITREMAAGLDEIIAKLT 943
 QY 860 ELBALQSKQKAEKLTRELKQKQKQVEIILREKIEDLQRMKEQO----- 906
 DB 944 EKKALQEAHQALDLOAEEDKVNLTLSKYKVLBOQVDDLESLQSKKVRMDLERAKK 1003
 QY 907 ---ELSTREASIQ-----KLOERDQELAR 928
 DB 1004 LEGDLKTQESIMDLENDKLQLEKLLKKKFPDISQONKIEDEQALALQKJLKENQAR 1063
 QY 929 LEE-----EACRAQOEFLSNFD--RIDECVNIERSLSG--SEFS 967
 DB 1064 IEELEELAEERTAKAKVEKLSRDLTRELIEISERLE--AGGATSVQIENKKGREAEFO 1121

```

QY 968 S---ELASACEKPNF-----NFSPYPREEVDGFEADDAA 1002
DB 1122 KMRDLBEATLQHATPAALRKHADVAELGQIDNLQVVKQLEKEKSEFLELDDVT 1181
QY 1003 FKSPNSSEHGHSDQRTSG-----IRTSDDSSSE-----DPTMNTVVPPTSBSAD 1047
DB 1182 -----SHMEQIIKAKANIEKVSRTLEDANEXRYKLEBAQSLNFTTQRAKLOT 1231
QY 1048 STVLLAPSVODSGSL--HNSSSGESTYCPQN-----AGDLPSPDGPDYD 1090
DB 1232 ENGEIARQLEKEKALIMQLRKLSYQOMEDLKQLEEGKAKNALAHALQARHDCUL 1291
QY 1091 DDDYEDGAIITSGS---SVTFMSYSGQSPDYRCV--GTNSSGAYRFSSGAGSSFE 1145
DB 1292 LRQYEEEMEKAKELQVLSKANSEVAQMRKYETDAIQRTTELEAKKLAQRLDA-B 1350
QY 1146 DSEEDPDRFDTDELSTYR-----RDSVYSCVTLYFHSFLYMKGLNWSK 1192
DB 1351 EAVEAVNAKCSLEKTYHRLQNEIEDLMVYERSNMAAALDKKQRF---DKILAEMK 1406
QY 1193 RRV-----CVLKDETFIMFRSKQBALKQGMILKK-----1221
DB 1407 QXVEEESQSELESSQKEARSLSTELFKYKAYEBSLEHLETFKXENKLOEISDLTBOQL 1466
QY 1222 GGGSTLSRRNMKKRWFVL---QSKL---MYFENDSEKLGTYEVRTAKEIIDNTTK 1274
DB 1467 EGKKNVHELEKIRKOLEVEKLELQSLAEBAASLEHEGKILBAQLEFNQIKAEIRKLA 1526
QY 1275 ENGIDIMADRTFHILAESPE-----DASQWFSVSQVHASND 1312
DB 1527 EKQEBEQAKRNHLRVVDSLOTSLDETGRNBLRYKKMGEDLMEIQLSQAARLVS 1586
QY 1313 QEIQEMHDEQANQNAVGLDVLIDVSCASDPDRPNSFVITITANRVLHCNADTEBEMH 1372
DB 1587 EAQKHLKNAQAHKADQLOLD---DAVRANDL-KENIAIVERRTLLQALB---EELR 1638
QY 1373 HMTLLORSKGDTRVBSQEFIVRGWLHKEVNSPKMSLSKLKRWPFVLTNLSIDYKSSS 1432
DB 1639 AVEEOTERSR---KLAEQELI-----ETSERVQLLHSQNT---SLINQKKMD 1680
QY 1433 KNALKLGTLVNLCSVPPDEKIFKETGYMNTVVGGRHGYLTKLNLNARMSVIO 1492
DB 1681 ADLSQLOTEVEEAVQECRNAEBA-----KKAITDAAMMAEBLK 1719
QY 1493 NVTDTKAPIDTPTQOLIQIDIKENCILNSDVVEQI 1525
DB 1720 KEQDTSAHLERMKKNMEQTIKDLQHRUDEARQI 1752

RESULT 5
US-08-905-223-395
; Sequence 395, Application US/08905223
; Patent No. 6223029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclercq, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223

```

```

; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 395:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 103 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: s1g_peptide
; LOCATION: -72..-1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 3.6
; US-08-905-223-395

Query Match 4.9%; Score 521; DB 4; Length 103;
Best local similarity 100.0%; Pred. No. 2.4e-33;
Matches 102; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 643 MPDQPDQAVNLQNLRYSGMLFTVIRKAGYAVRPPDFFKRYKYLMMNLALPEDVRGKC 702
DB 1 MPDQPDQAVNLQNLRYSGMLFTVIRKAGYAVRPPDFFKRYKYLMMNLALPEDVRGKC 60
QY 703 TSLDLYDASNSEWOLGKTQVFLRESLRQKLEKREBEVSHA 744
DB 61 TSLDLYDASNSEWOLGKTQVFLRESLRQKLEKREBEVSHA 102

RESULT 6
US-08-353-700-1
; Sequence 1, Application US/08353700
; Patent No. 5599919
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATTNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO: 1:

```

SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: HUMAN
US-08-353-700-1

Query Match 2.5%; Score 265.5; DB 1; Length 3248;
Best Local Similarity 18.4%; Pred. No. 6,4e-11;
Matches 422; Conservative 362; Mismatches 764; Indels 741; Gaps 110;

136 ECVRCMLKHNOCILIKSGSAGKTESTLIKFLSVISQSLSLSEKTSCEVERALI 195
196 ESSPIMEAFGNATVYNNNSRFGKFLVQNIQKGNIG--GRIVDCILSSQ-----NR 247
609 -----YELKEKTLSCSKSENEKILLTOMSEKEMLOSKINLETCLTKQILKSHRYN 663
248 VVRQNGE-----RVYH-----IFVALLAGLEHEEREEF-----YL 278
664 RVYTLTMDRENLSVEIRNLHNLVDSKSVETOKLAVMLOKAFESDQKHQEIENMCL 723
279 STE-----NYHLNOSGCVEDKTISDQSFREVTAMVMQSKVEYE 323
724 KTSQLTQVDELEHLQOLISNEIMDKRCYQD-LHAEYESLRDLKSKASLVTNEDHOK 782
324 VSRLLAGLILHNIETITAGAGAVSEFKALGRSAELGLDPTOLTALTQSRWFLGEEI 383
783 --SLLA-----FDQPRAMHSAFNTIGEGSM-----PSRSECRLEADSPKASAI 827
384 LTPV--NVQOAVDSRDSLAVLYACCFEYVYIKKINSRINKNEDFKSIGILIDIFGFENFEV 441
828 LQRRVDSLEPFSLEKQMSNDLOKCEELV-----QIKGEIE-----EN--L 867
442 NHEQGNINYANKLOEYFPAKHFPLEOLEYREGVWEDIDINDGECDLIEKLGILL 501
868 MKAEQMSQSFVATSQ-----RISKLOEDTSAHQNIV-----AETLSALENEKEEL 913
502 ALINESHPQATDSTLTLEKLSQHANNHFYKPRVAVNVFVGHYAGVGVVGRIGILEX 561
914 QILNDKVEYEQ-----BIOE-----LKK 932
562 NRDTFRDDLNTL-----LRESRDFIYD-----FEHVSRNODPTLKGSKHRRPTV 609
933 SNHLBDSLKEQLQLSETLSLEKEMSSISLNKRBEIHLTQENG--TLK-----EI 982
610 SSQPKVDSLHSLMATISSNPPFVRCIKENMQMPQFQAVVLANQSRISGMLETYRIRK 669
983 NSLINOBNKMLIKSSSFAN--YIDERKESISLSLOYIOEKLI-----LIQ--RCEB 1031
670 AGAVNRPDPQFYKYVLMRNLALPDEVGKCTSLLOLYDASNSEMOGKTAVFLRESI 729
1032 TGAH-----YEDLSQTKAAQEKNSKLECLNCTSLCE-----NRKNEL-----EQ 1074
730 EQLERKREBEVSHAAMVIRAHVIGFLARKQYKVLVCVYIIOKNYRAPL----- 779
1075 KEAFALKEHOFELKLA-----FAEBRNQNLWLEJETVOQALRSEMTDQNNKSKE 1124
780 ---LRRFLMLKKAALVFOQLAGQIARVYRQLLAEKVEBEKKKQEBEKKREBE- 835
1125 AGGLKQEIIMTLKEQNMQKEVND-----LQENEQMLKVMKTKIECNLTSEPT 1174
836 -----REERERREARALAOOEETRKOQ-----ELBALQSKQEAFLTELEKOK 882
1175 RNSVKERESERNQCNFPQNDLVEKLSLDSYNAQLVQLEAMLRN--KELKL--QESERKE 1232
883 NKQVE--EILRLKEIEDLQRMKEQ-----ELSLTEASLQRLQER 921

1233 CLQHEIQTIRGDLTSLNLODMQSOBISGLKDCEDIAEBEKYISGRPHLSTSQDNHNLQCS 1292
922 RQGEIIRLE--BEACRAQ-----EFLSELNPEIDCYR-----NI 956
1293 LQTTMKNLNELEKTEIILDAEKYELVTELN--DSRSECTITATKMAEYVCKLNEKYILND 1351
957 ERSLSGSEFSSELASACEEKN-----FNFSQPY-----PIREVDGEFPAADDA 1002
1352 DSGLLHG--ELVEDITPGEGGEGQPNQHPVSLARLDESNSYBHLTSLDKVQNHFLQELK 1410
1003 FKDSNP-----SEHSDQRTSGIRTSDDSEEDPYRNMOT----- 1038
1411 FLSQSEHKILHDQCMSSKMSSELQTYVDSIKKANLVLSYLNRNFGDLVKEMOLGLEE 1470
1039 --VPTSPASDSTVLLAPSVODSGSLHNSSGSESTY--CMPQNAQD---LPSPDGYVDQ 1092
1471 GLVPS-----LSSSCVPDSSSL--SLGDSSTYRALLBOTGMSLISNLEGAVSANQ 1520
1093 DDYED-----GAITSQSVTFPSN-----SYGQMSPDY 1120
1521 CSVDVFPSSLOTYVDSLKAENLVLSYLNRNQGLVKEMQGLEGLVPSLSSSCVPD- 1579
1121 RCSVGTNSSGAYR-----FSSGAQSSFEDS-----BEDFSRFTD- 1158
1580 SSSLSSIGDSSFYRALLBQTDGMSLISNLEGAVSANQCSVDVFPSSLQEBNLTREKTPS 1639
1159 -----DELS-----YRRDSVYSCVTLPYHSLFLMKGLMN--SKRRKRCVLDKTEFLMF 1206
1640 APAKVEEIESLCEVYRQ-----SLEKLEKMSQGIKMKKEIOBLEQLLSSB----- 1687
1207 RSKOBALQGMWLHKKGSGSSTLSRRNMKRMFVLQSKLMPENDSEKL-----KGTAV 1262
1688 ROEDLCRKQYLS-----NBQW-----QQLUTSVTLMESLTALEKKQTEBL 1730
1263 RTAKEIIDNTKENGIDIMADRTHLIAESPEDASQMFV--LSQVHASTDQEIQEMH 1319
1731 SLELEVA--RLQLOGLD--LSGRS--LIGIDTEDAIQGNESCDISKEHSTETTERPAG 1784
1320 D--EQANPONAVG--TLVDGLIDSV-----CAGD--SPDR----- 1348
1785 DVHQICDQAQODNLIDIEKITETGAVKPTGCSGQSPDTNYBPFGEKQSGSECSISE 1844
1349 -----PNSFVL-----TAN-----FVLHCNADTFEEMHMTLLQSRK 1382
1845 LSFSGBNALVDPDFLQNGEDINLQRYKEISNEMLRLHVLVEDRKYVE--SLINEMK 1901
1408 MSS-----LKLKKW-----FVLTHNSLIDYKSSFKNALKGLTVLANS 1446
1962 LNSDEMEADKSSREDIGDGNVAKVMSKERFLDVENEISLRIR--SEKASIEBHAYLEAD 2020
1447 CSVVPDEKIFKGTQGMVNTVYGRGVCYRLVYKLINEATBMSVQVONTVDTAPIDTPQ 1506
2021 LEVY--QTEKLCLEKNEN-----KQVYVCEELISVYTSBNQRLGELDTNSKTTALD 2074
1507 QLIODICE-----NCINS--DVVEQIYKRNPIRLYTHNPLHSPLLPLRYGDIINL 1555
2075 QLEKMKKEKTOLESQESCHICIQVAEAYEK-----TBLQTLSSDVS--EL 2122
1556 LKQGYTTLQDBALIKFISNLOLESMSDPIPIQIGILTQGHDLRLRDBELVQILKQYTK 1615
2123 LKDX--THLQEK-----LOSLEKDSQALSLTKCELB--NOIADLNKEKL--LYKES 2170
1616 VHPGSVGNLYSMOITCISCFPLPSRGLIKYKFLKRIROFQGTEN-----EKY 1667
2171 LQARLSBQYELQANYSKALBALVYKGFALRLSTQSEVYNQLRGIRIKLVRIDAEKX 2230
1668 ALFTYESLKTQCRFPVSRDEIALIHRQEWTSYVYCHGGGSCKITINSHTAGEV--- 1724

Db 2231 QLIHAEKLERE-RENDLSLKQVENIEREQMSEB-----NOELVILDAENSKAVEYL 2283
Qy 1725 ---VEKLRIGLAMEDSNMFPALFEYNGHVDAIESR-----TVADVLAKEFKLAATSEVG 1777
Db 2284 KTOIEBARSJKLFE-LDVLTLSEKENLTKOIOEKOGULSELDKLSKSLLEKEQA 2242
Qy 1778 DLFMKFYFKLYCLDLDNVKDSVEFAFMFEQAHEAVIGHHAPPEENLOVIALRLQYL 1837
Db 2343 EIQIK-----EESKTAIVEMLOQNLKEINAVAA-----L 2371
Qy 1838 QGYTLHAAT-----PLEEYV-----SLQRLKARISSTKTFPCCELEKRRRSFLEGL 1888
Db 2372 CGOQETIKATQGLDPRIEEHQLRNISTEKLARL-----EADKKQLCVLQ-QL 2420
Qy 1889 RRSFRTGSVVRQKVEEQMLDMWIKEEVSARASIIDKMRKFOQMNQOQAMAKYMALIKE 1948
Db 2421 KESEHNADLKGVEN-----LERLEIART-----NOEHALE-----AEN 2457
Qy 1949 WPGYGTLLFDVECKEGFPQELMIGVSADAVSVYKGGEPLEFVFOYEHILSFGAPLANT 2008
Db 2458 SKGEVEYL---KAKIEGMTQSL-RGLELDVVTIRSEKENLTNELQKEQERISLEIINSS 2513
Qy 2009 YKIVUDERE 2017
Db 2514 FENILQKE 2522

RESULT 7
PCT-US95-16216-1
Sequence 1, Application PC/TUS9516216
GENERAL INFORMATION:
APPLICANT: Yen, Timothy J.
APPLICANT: Ractner, Jerome B.
TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dann, Dorfman, Herrell and Skillman
STREET: 1601 Market Street Suite 720
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103-2307
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/16216
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/353,700
FILING DATE: 09-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Reed, Janet E.
REGISTRATION NUMBER: 36,252
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 563-4100
TELEFAX: (215) 563-4044
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3248 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
HYPOTHEtical: NO
ANTI-SENSE: NO
PCT-US95-16216-1

Query Match

2.5%; Score 265.5; DB 5; Length 3248;

Best Local Similarity 18.4%; Pred. No. 6.4e-11;
Matches 422; Conservative 362; Mismatches 764; Indels 741; Gaps 110;

Qy 136 ECVRCIMKNDHQCLIKESGAGTSEFKLLKFLSVISQOSSLSELSKEKSCVERAIL 195
Db 568 DCSQDLKREHIEQLNDLKSSTKESKAL-----SALBLKKE----- 608
Qy 196 ESSPIWEAFNAGKATVYNNNSRFGKLVOLNCOKNIOG--GRIVCILSSQ-----NR 247
Db 609 -----YEBLKEKERTLPSCKSKSENEKLLTQMESEKENLOKINHLETKLQOIKSHEVNE 663
Qy 248 VVRONPGE-----RNVH-----IFALLAGLEHERREEF-----YL 278
Db 664 RVTLEMEKRENLVSEIRNLHNLVLDKSEVEETQKLAHYELOQKAEFSDQKQKEIENMCL 723
Qy 279 STPE-----NYHYLNOSGCVEDKTIISDQSPREVIATDVMOFSKEEVE 323
Db 724 KTSQLTQGVEDLEHKLQILSINIMDRCTYD-LHAETESLIDLKSDASLVNEDHOR 782
Qy 324 VSRLLAGLILGNIEFITAGAOVSFKTALGRSAELGLDPTQLTALTORSMFLRGEI 383
Db 783 ---SLLA-----FDQPAWHHSFANILIGQSM-----PESRECRLEADQSPKNSAI 827
Qy 384 LTPL--NVOQAVDSRDSLAMALYACCFEYVYIKINSRIKGNEDFSGIGLIDIFGENFEV 441
Db 828 LONRVDSLEFSLESOKMNSDLOKOCERLY-----QIKGEI-----EN--L 867
Qy 442 NHPEOFNINVANKOEYFNKHIIFSLQLEYSREGVWEDIMDNGECLDLEKKGL 501
Db 868 MABQWHSFVAFETQ-----RISKLEBTSANOV-----AETLSLENKEKEL 913
Qy 502 ALINESHPPOATDSTLEKLSHONHNFYKPVAVNPFVKGAYAGEVQYDVGILK 561
Db 914 QLNLDKVEREQ-----EIQ-----LKK 932
Qy 562 NDTFRDILNL-----LRESRFDLYD-----FEVSRNNODTLKCSKRRP 609
Db 933 SNHLEDSLEKELQILSETLSLEKKEMSSITSNKREIELTQENG--TLK-----BI 982
Qy 610 SSQFKVDSLHSLMATLSSNPFVRCIKPMOKMPOFOQAVNLQRLRSGMLFTVRIRK 669
Db 983 NASLNOEKMLIQKSEFPAN--YIDEREKISIELSDQYKQKELI-----LQO--RCE 1031
Qy 670 AGYAVRRPQDFYKRYKVLNMLALBEDVGRKTSILQLYDASNSEWQGLKTVLRSL 729
Db 1032 TQNA-----YEDLSQYKAAQENKSLCECLNCTSLCE-----NKNEL-----EGL 1074
Qy 730 EQLKREKREBEVSHAAVIRAHVGLPLARKQYRKVLVCVVIQKRYRAF----- 779
Db 1075 KEAFKKEHOFPLTKLA-----FAERRQNMLLELETVOQALRSMTDNONNSKE 1124
Qy 780 ---LRRFLHLKKAIVPQKOLRGQIARVYNQALLAEKREBEKKOEKKEEKEE 835
Db 1125 AGGLQOETLWLEKEDONKQKQEVND-----LQENBQMLKVMKTHNCOINSEBET 1174
Qy 836 ---REREREREAELRAQOEETRQO-----ELEALQSKQKAEMLTRELEKQKE 882
Db 1175 RNSVVERESERNQCFKQMDLEVAKEIGDSYNAQLVQLEMLRN--KEKL--QSEKKE 1232
Qy 883 NKQVE-EILRLEKEIEDLQKMEQO-----ELSITEASLOKQER 921
Db 1233 CLQHELOTIRGDLSTNSLQDMOSQETISGLKCEIDAEBKYISGPHLSTQNDNAHLOS 1292
Qy 922 RQGLRRLLE--BEACRAQO-----ELESINPDEIDCYR-----NI 956
Db 1293 LOTTNKLNELEKICETIIQAEKYLELTVELN--DSRSECTIATRKAAEYKGLNEVKILND 1351
Qy 957 ERSLSGSEFSSELASACEEKP-----FNFSQPY-----PEEVDGEFADDA 1002
Db 1352 DSGLLHG--ELVEDIDGEGFQEPQNHQVSLAPLDESNSYHLLTSDKEVQNHFLQEK 1410
Qy 1003 FQDSFNP-----SEGHSDQRTSGIRTSDDSEEDPYANDT----- 1038

1411 PLSLOSEHKLILHDQHCQSSKSELOQYVDSLKAENLVLTNIRNFQGLVEMQGLSE 1470
1039 -VPTSPASDSTVLAPSVODSGSLNHSSESTY--CMFQONAGD--LPSPDGDYDQ 1092
1471 GLVPS-----LSSSCVPDSSSL--SLIGDSSFYRALLBQDGLSLNLEGAVSANO 1520
1093 DYED-----GATSGSSVTPSN-----SYGQSPY 1120
1521 CSDVEFCSSLQTYVDSLKAENLVLTNIRNFQGLVEMQGLSEGLVPSLSSCVPD- 1579
1121 RGSVGTYNSSGAVR-----FSSGQSSPDS-----BEDPSSRDTD- 1158
1580 SSSLSLSDSSFYRALLBQDGLSLNLEGAVSANOCSVDVEFCSSLOEENLTREKTPS 1639
1159 -----DELS-----YRDSVYSCVTLFPHSFLYKGGIMN-SMKRMCVLKDETFMFW 1206
1640 ARAKVEELESCEVTRQ-----SLEKLEKMEGQGINKEIOLEOLLSSE----- 1687
1207 RSKQKALKQGLHKKGGSSSTLSRRMKRPFVLRQSKLWYFENDEEKL---KGTVEV 1262
1688 RQELDLCKQYLSE-----NEQM-----OQKLSVTLWESKLAERKQTEOL 1730
1263 RRAKEIIDTTKENGDIIMADTFHLIASPEDASQWRSV--LSQVASTDQEQENH 1319
1731 SUELEVA--RLQQLGLD--LSSRS--LLGIDTDAIQGRUESCDISKHTSETTERTPKH 1784
1320 D-EQANPQNAV--TLVDGLIDSV-----CASD-SPDR----- 1348
1785 DVHQICDKQADQDLNIDIEKITETGAVKPRGEGSGSPDTPYPERGEDTKQSSSECISE 1844
1349 -----PNSFVIT-----TAN--RVLCNADTPEEMHMTLLQRSK 1382
1845 LSFSGNALVPMDFLQNOEDIHNLQRLVKETSNEMLRLVLRDRKRYE--SLINEMK 1901
1383 G-DTRBEGEP-----YRGMLHKE-----VQNSK 1407
1902 ELDSKLHLEVOJMTKIEACIELEKIVGELKENSPLSELEVPSCDHQELLQRYVETSEG 1961
1408 MSS-----LTKKKRM--FVLTHNSLDYVYKSEKNAKGLTYLVNSL 1446
1962 LNSDLEMAHDKSREDIGDNVAKVNDWKRFLDVENELSRIT-SEKASIEHVALYLEAD 2020
1447 CSVPPDEKIFKETGYMNTVYVGRKCYRLYTKLNEATRMSSVIONVDTKAPIDTPRO 1506
2021 LEVV-QTEKLCLEKONEN-----KQVYVCLSEELSVVTSERQNLGELDTMSKTTALD 2074
1507 QLIQDIKE-----NCLNS-DVVEQTYKKNPILRTYHPLHSPPLPLPYGDINLNL 1555
2075 QLSERKKEKTOBLESHQSECLHCIOVAEAEVKEK-----TELIQTLSSDVS-EL 2122
1556 LMDKGTTLQDEAIKIFNSIQLESMSDPIPIQIGLQTHDLRLADELYQQLIKQNTK 1615
2123 LMDK--THLOEK-----LOSLEKSOALSITKCELE--NOQAQNKKEEL-LVRESSES 2170
1616 VPHPGSVGLYSGWQILTCLSCFLPSRGILKYLKFLKRIEOPGTEM-----EKY 1667
2171 LQARLESSEYBKLAVSKALAEALVKEGEPALRSSTQEBVHQGRGIEKLARVIREADEKK 2230
1668 ALPTYSLKTKREVPSRDEIHALIHROEMSTYVCHGSGCKTINSHTTAGV--- 1724
2231 QHIAIAKIKERE--RENDSLKDKVENLERLOMSEE-----NOELVILAENSKAVEVTL 2283
1725 -----VEKLIRGLAMEDSRMFPALFVNGVDAKIESR-----TVADVIAKPEKLAASENG 1777
2284 KIOIEMASLKTFF--LDLVTLSEKENLTKQIQERQOGLSELDKLSSFKSLLEKEQA 2342
1778 DLPWKEVFKLYCFLDTPNVKQSVBPAFMEOAHEAVIIGHNPAPEENLOVLAALRQVYL 1837
2343 EIQIK-----EESKTAVENTMQLQNTKELNEAVNA-----L 2371
1838 QGDTYTHAAI-----PPELEVY-----SLQRLKARISOSTKTPPCERLEKRTSPLEGLT 1888
2372 CGQOETMKATBOSLDPPIEEHQLRNSIRKLRARL-----BADEKKQCLVIO-Q 2420

1889 RRSFRTGSVROKVEEQWLDMMIKEEVSARASIIDKRRKEQGNQEQAMAKYNALIKE 1948
2421 KSEHNAHDLKGRVEN-----LERLEIARNT-----NQERAALE-----AEN 2457
1949 WPGYSTLPDVECKEGEPFOELMGLVADAVSVYRGRGEPLEVPQYHILSFGAPLANT 2008
2458 SKGEVETL---KAKIEGHTQSL-RGLBIDVVTIRSEKENLTJELQKQERISELBIINS 2513
2009 YKIVDERE 2017
2514 FENILQEK 2522

RESULT 8
US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022
GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores
STREET: 4370 La Jolla Village Drive, Suite 700
CITY: San Diego
STATE: California
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/328,254
FILING DATE: 24-OCT-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/141,239
FILING DATE: 22-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Campbell, Cathryn A.
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 1191
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 535-9001
TELEFAX: (619) 535-8949
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 2482 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-328-254-6
Query Match 2.5%; Score 264; DB 1; Length 2482;
Best Local Similarity 18.4%; Pred. No. 5,4e-11;
Matches 401; Conservative 339; Mismatches 729; Indels 710; Gaps 102;
135 NECYRCLMKRHNQICILKNG-----ESGAGKTESFKLILFSLVISQOSLSLKEKTS 189
30 NERVATTEMDRENLSVEIRNLNVLNVDSSVEVETKLAEMEL-----OQKAEFSQKQK 85
190 VERAILLESSPIWEAFGNKATVYNNNSRFGKVFQUNICQKGNIOGRIYDCLSSQNRV 249
86 IE-----NMCLKTSQLNGQYED----- 102
250 RQNPGRNTHIFYALLAGLEHREKREFFLSTPENRHYLNQSGCVAEKTIISQDESFRVIT 309
103 -----LEHK-----LOLLSN-EIMDKORCYOD-LHAEYSLRDLK 136

QY 310 AMDVMSKEEVBRESVLLAGLILHGNIEFTAGQAVSFKALGSAHLLGDTPTQTD 369
 Db 137 SKDASLVATNEDHQR--SLLA-----FDQOPAMHNSPANTIGEGSM-----PESRSE 181
 QY 370 ALTORSHFLGEEILFTL--NVQOAVDSRDSLALALYACFEVWIKINSRIGNEPDFS 427
 Db 182 CRLEADQSPKSAILOVRVDSLEFSLESOKMNSDLOKOCLELY-----QIGGEIE-- 232
 QY 428 IGLIDIFGFNFVEVNEHGFNINAVANEKLOEFNKHIFSLQOLEYSNEGVLWEDIMID 487
 Db 233 -----EN--LMKAEQWHSFVAETSC-----RISKLOEOTSANQV----- 267
 QY 488 GEGLDLEKKLGLLALINEESHPPQATSTLEKLSQHANNHFYKPRVAVANNFGVKYH 547
 Db 268 AETLSALENKEKELOLNDKVETEQA----- 233
 QY 548 AGEVQDVVRGILEKNRTPFDLLNL-----LRESRDPITYDL-----FEHVSRRNO 595
 Db 294 --BIOE-----LKKSNHLEBDSLEKLOLSETLSEKEMSSIIISNKRIBELTJONG- 345
 QY 596 DTLKCSKRRPVTYSQFVDSLHSLMATLSSSNPFVRCIKPMQOMPOFOQAVVLAQ 655
 Db 346 -TLK-----EINASLNEKKNLLOKSSFPAN--YIDEREKISIELSDQYOEKULI-- 392
 QY 656 LRYSGMLETYRIRKAGYAVRRPODFYKRYKVLMRMLALPEDVRCGKCTSLLOLYDASNS 715
 Db 393 -----LLQ--RCSETGNA-----YEDLSQKYKAAQENKSKLECLINETSICE-----NKR 436
 QY 716 WQUGTKVFLRESLEQLEKREBEVSHAMVIRAHVGLFARKQYRKVLVCVITIQKYN 775
 Db 437 NEL-----EOLKEFAKEHOFITKLA-----FAEERNQNLMLBELTVOAL 478
 QY 776 RAFL-----LRRFLHLKKAALVIFOKOLRGQARVYQQLAEKREOEKK 822
 Db 479 RSEMTDNONNSKESAGLKOELMTLKEONKMOKEVND-----LLOEBOUKMVA 528
 QY 823 QEEBEKKREBE-----REREREAREALRAOEETBTRKO-----ELEALQSQ 868
 Db 529 KTHGECONLESEPIRNSVKESEBQCNFKPQMDLVEKISLDSYNAQVQLEAMLRN- 587
 QY 869 KEAELTRELKOKENKQVE--ELIRLEKEIEDLOPMKEQ-----E 907
 Db 588 KEELT--QESKEKECLOHELOTIRGDLSTSLQDMOSQOELSGKDCIDEAEKYISGPH 646
 QY 908 LSLTEASLOKLOERRODELRLLE--BEACRAAO-----EFLSELNPEIDECVR----- 954
 Db 647 LSTSONDNALQCSLOTTMKNLEKICEILOAEKELVTEIN--DSRSCITATKMAE 705
 QY 955 -----NIESLSGSESEFSELASACEEKN-----FNFSQPY-----P 988
 Db 706 EVGKLINLVKILNDGSLHIG--ELVEDIPGEGFGEQNEQHPVSLAPLDESNGYEHULTS 764
 QY 989 EEEVDGEFEADDAPKDSNP-----SEHGSDORTSGIRTSDDSEEDRYMNDT----- 1038
 Db 765 DKEVQMHFAELOKFLSLQSEHKILHDQOMSSKSELOTYVDSIKAEMLVISTYLIRNF 824
 QY 1039 -----VPTSPASDSTVLAPSVODSGSLHNSSGSTY--CMPONAGD- 1080
 Db 825 QGGLVEMQGLEGLVPS-----LSSSCVPDSSSL--SSLGSSFPRLALLETGTGM 874
 QY 1081 --LPSPDGDYDQDDYDEGALITSGSVTFPSNYSYGSSOWSPDYRCSVGTYNSSGAYRFS- 1137
 Db 875 SLTSLNEGAVSANQCSVDEVFSGSLQOEENLT-----RKETPSAPAKVELESEI 923
 QY 1138 -EAGQSSFEDSEEDPDSRFPDTDELSTYRDSVYSCVTLPYFHSFLYMKGLNMSMKRMC 1196
 Db 924 CEVYRQSLKLEKEMESQ-----GIM--TKKEI 949
 QY 1197 VLDEFTFLWRSGOALKQGLHKKGGSSSTLSRRMKKRWFLRQSKLWYFENDSEBK 1256
 Db 950 QELBQLLSERQELDLCKQYLS-----NEQW-----OQKLTSTVLEHESKL 992
 QY 1257 ---KGTVEVTRAKEIINDTTKENGDIIDWADRTFHILAESPDASQWFSV--LSQVHA 1309

Db 993 AAEEKQTEQLSLELEVA--RLQLOGID--LSSRS--LLGIDTEDALQGENESCDSIKERT 1046
 QY 1310 STDQELQEWND--EQANPQNAV--TLVDGLIDSV-----CASD--SPDR----- 1348
 Db 1047 SETTERPRGHVHQICDDAQODLNDLIEKITETGALKPTGECSSQSPDTNYEPPEDEK 1106
 QY 1349 -----PNSFVIL-----TAN--RVLHCNADTPEEMH 1372
 Db 1107 TQGSSECSIELSPGPNALVPMDFLQNGEDIHNLQRAVETSENEMRLHLVIEDRKYE 1166
 QY 1373 HWITLQSKG--DTRVEGEF-----IVRGWLHKE----- 1401
 Db 1167 ---SLINKEKEIDSKHLQEOVLMTKIEACIELEKIVGELKEMENDSEKLEFSCDHO 1223
 QY 1402 -----VKNSPKSS-----LKLKRR--FVLHNSLDYKSSBKAL 1436
 Db 1224 LLQVETSEGLNSDLEMTADKSSREDIGDNVAKVDSMKERFLDVENELSRIR--SEKAST 1282
 QY 1437 KLGLVLNLSGVPPDEKIFKETGYMNVTVYGRKCYRLYKTLNEATRWSSVIONVTD 1496
 Db 1283 EHEALYLEADLEV--QTEKLCLEKNEN-----KQKIVCLEEELSVTSEBQRLGELD 1336
 QY 1497 TPAPIDTPQOLIDIKE-----NCLNS--DVEQIYKRPDILRYTHHPLHSLIP 1545
 Db 1337 TWSKXTTALDQSEKKEKTEQLESHQSECLHICIOVAEAVEREK-----TELLQ 1385
 QY 1546 LRYGIDINLNDKQYTIQDQAIKIFNSLOQESMSDPIPIQGLIQTHGLRLPLRDL 1605
 Db 1386 TLSSVDS--ELLKDK--THLOEK-----LOSLEDQSLSTTKCELE--NOIADLNK 1433
 QY 1606 YCOLIKQTKVPHPSVGNLYSMQILCTLSCFLPSRGLIKYKHLKRIREQFGTEW- 1664
 Db 1434 EL-LYKESSELOARISESDYEKLANSKALEALVKEGPALRLSSTQSEVHQLRGIEKL 1492
 QY 1665 -----EKVALFTYESLKTCKREFVBSRDEIHALHROENTSTVYCHGGSCKITTS 1717
 Db 1493 RVRIBADEKQOLHIAEKLKERE--RENDSLKQVEMLERELQWSE-----NOELVILDA 1545
 QY 1718 HTTAGEV-----VAKLRGLAMEDSRMFLFEVNGVYDKAIEER--TVADVLKFE 1767
 Db 1546 ENSKAEVETLQTOIEMASLKVFE--LDLVYTRSEKENTQIOERQOGLSLDKLSSF 1604
 QY 1768 EKLAAITSEVGDLPMKFYFKLFCFLDTDVVPKQDSVFAMFEOAHAVIHGHHPAREENQ 1827
 Db 1605 KSLBEKEBAEIQIK-----BESKTAVENTLQNLKEINBA 1639
 QY 1828 VLAALRLOYLOGDYTLHAI-----PLEBEVY--SLQRLKARISQSTKFTTPEERLEK 1878
 Db 1640 VAA-----LCGDQOIMKATEQSLDPPIEBEHQLNLSIEKLRARL-----EADBK 1683
 QY 1879 RRTSFLEGTLRSPFTGSVVRKXVEBQMDMWIKEYSSAASIIDKMRKQGNQOBA 1938
 Db 1684 KQLCVLQ--OLKSEHHADLKGIVEN-----LERELEIAT-----NOEBA 1723
 QY 1939 MAKYVALIKEMPGYSTLFEDVECKEGGFQOELMLVSDAVSVYKRGEPPLVEFYQEH 1998
 Db 1724 ALE---AENSGEVETL---KAKIEGNTQSL--RELELDVYTRIBEKENTLNELOKEQER 1775
 QY 1999 ISFGAPLANTYKIVDERE 2017
 Db 1776 ISELINSSFENILOEKE 1794

RESULT 9
 US-08-929-329-5
 ; Sequence 5, Application US/08929329
 ; Patent No. 6120770
 ; GENERAL INFORMATION:
 ; APPLICANT: Adams, John H
 ; APPLICANT: Dalton, John P
 ; APPLICANT: Kappe, Stefan
 ; TITLE OF INVENTION: Plasmodium Proteins Useful for Preparing

```
/ TITLE OF INVENTION: Vaccine Compositions
/ NUMBER OF SEQUENCES: 23
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Barnes & Thornburg
/ STREET: 11 S Meridian
/ CITY: Indianapolis
/ STATE: Indiana
/ COUNTRY: USA
/ ZIP: 46204
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/929,329
/ FILING DATE:
/ CLASSIFICATION: 424
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Breen, John P
/ REGISTRATION NUMBER: 38, 833
/ REFERENCE/DOCKET NUMBER: 835910-28685
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (317) 231-7745
/ TELEFAX: (317) 231-7433
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1507 amino acids
/ TYPE: amino acid
/ STRANDEDNESS:
/ TOPOLOGY: unknown
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: Plasmodium yoelii
/ US-08-929-329-5

Query Match      2.4% Score 255; DB 3; Length 1507;
Best Local Similarity 17.6%; Pred No. 1.2e-10;
Matches 164; Conservative 159; Mismatches 281; Indels 326; Gaps 35;

QY 441 VNHFOFNINVANKEKLEQEFNKHIFSLLEOLEYSRBLVEMID-----WIDNG----- 488
DB 673 INGNHLNMFLEIRGTGLNKMKNVLEPYEKSLAMHGDLSVCPSWDEBNLYKKN 732
QY 489 -----ECIDLIEKKL-----GL----- 500
DB 733 RDYNDMCKSTWKSSTIPLKMPDYKTKLLYGLYGLGRLGSLNISKVNIKFSQPNNT 792
QY 501 LALINEES-----HPPQATDSTLLEKLSHGHAN 528
DB 793 LPMFNPSSIKNLLDLSVYCLGPLENAYNNKCFRSLFAVFNHETNECIIIGTHEQERN 852
QY 529 NH-----FYVXP-----RVANNFGVGHYAG 549
DB 853 NNCRTRRSDTDKPNQNVAKNISTKNWTVVTSFIRDYEEKCPRPYPLKFKSFG----- 906
QY 550 EYQYD-----VAGILEKRDTRDILNL-LRESRDFIYDLFEHVSNN----- 593
DB 907 --KYDEETGCKSLINK-----NINIPLFSSCLEVNFIMYPSVLTQTEKKNYGWV 958
QY 594 -----NQDTL-----KCGSKHRPTVSQPKVDSL-HSLMATLSSNPFVRCIKPNMK 642
DB 959 ASESVNSSLVYAKGECCYINEKNC-----VIDKVNHRFTSLTINDIDFNONI--NLVK 1012
QY 643 MBD-----QPDQAVVLNQLRYSGMLFTRIRKAGVAVRRPQDFYKRYKVMRLALPEDVR 699
DB 1013 LDELVINNDQSSSHNRKAKNTPIEN-----SESTIVRKNSNAPEHR 1054
QY 700 GKCTSLQLYDASNSWQGLKTVFRESLEQLEK--HREBVSHANAVIRAHVIGFLA 757
DB 1055 ---SLKINSYTRNGENPAKESDSTRNTDESIMDEVINKREBAKNABIRKK----- 1105
```

```
QY 758 RKOYKVLVCIWIIOKNYRAFLR-----RRFLHKAAIVFOKQFGQ-IARRVROL 811
DB 1106 -EEAKAMAKAEEERKKAELAVKAAEEERKRIEAKKAAEEERKRIEAKKAAEEERKRIE 1164
QY 812 AKREOEKK-----KOEERKKEERERERERERERELPAQOEETRKQOELALOKS 867
DB 1165 AEKKAEEERKRIEAKKAAEEERKRIEAKKAAEEERKRIEAKKAAEEERKKAELAVKAAE 1224
QY 866 OKAEETLELEKOKEN-----KOEELILEKEIEPLQMKQOELSL 910
DB 1225 KKAAPAKKALKESSRQKKTIEAVKAAEEERKRIEAKKAAEEERKRIEAKKAAEEERKRIE 1284
QY 911 TEASLOKLERRDQELRLLEERACRAQEFLESINFDIEDECVRNIERSLSGSEFSSSEL 970
DB 1285 VEA-----EKKAEEERKRIEAKKAAEEERIKR-----EAVKKEEVEVINKSNLSEK 1331
QY 971 AESACEKPNFNFSPQPYEEVEVDSGEFADDDAFKDSPPNPEHGSDDQT-----SG 1021
DB 1332 KTAIFLKKNNNY-----ETRNIDNDSFKLDEEBYKSRNIDYTRNKIISMSKEN 1381
QY 1022 IRTSDSEBDBPYNDTVPTSPSADSTVLLAPSVQDSGLNHSSSGSESTYCPQNAQDL 1081
DB 1382 MCTNDVSSKYCDYMKDKISSGCSNDEKQJCCSISD-----YCL----- 1421
QY 1082 PSPDDYDYPDDYEDGATSGSSVTFPNSYGSQWSPPYRGSVGTYNSSGAYRFSSEG-- 1139
DB 1422 ---NYPDNNSKTYD-----CTKKEFSD-----PLYKFSNBEYKAVY-FAGAGII 1464
QY 1140 -----AQSSEDESEDPD 1152
DB 1465 MSIIAICIKIIGKKMFKEVAFDEIVEDYD 1494
```

```
RESULT 10
US-08-728-323A-2
/ Sequence 2, Application US/08728323A
/ Patent No. 5948676
/ GENERAL INFORMATION:
/ APPLICANT: Chang, Yuan
/ APPLICANT: Bohenzky, Roy A.
/ APPLICANT: Russo, James J.
/ APPLICANT: Edelman, Isidore S.
/ APPLICANT: Moore, Patrick S.
/ TITLE OF INVENTION: Immediate Early Protein From Kaposi's
/ TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
/ TITLE OF INVENTION: Encoding Same And Uses Thereof
/ NUMBER OF SEQUENCES: 21
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Cooper & Dunham LLP
/ STREET: 1185 Avenue of the Americas
/ CITY: New York
/ STATE: New York
/ COUNTRY: U.S.A.
/ ZIP: 10036
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/728,323A
/ FILING DATE:
/ CLASSIFICATION: 435
/ ATTORNEY/AGENT INFORMATION:
/ NAME: White, John P.
/ REGISTRATION NUMBER: 28, 678
/ REFERENCE/DOCKET NUMBER: 0575/52268/CPW/MSC/SKS
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 212-278-0400
/ TELEFAX: 212-391-0525
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
```


QY 923 DQELRLREE 932
: : : : :
Db 441 EQLRREOE 450

RESULT 13

US-08-800-644-94
; Sequence 94, Application US/08800644
; Patent No. 5958752

GENERAL INFORMATION:

APPLICANT: Steinert, Peter M.
APPLICANT: Lee, Seung-Chul
APPLICANT: Kim, In-Gyu
APPLICANT: Chung, Soo-Il
APPLICANT: Park, Sang-Chul
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
TITLE OF INVENTION: Methods of Using Same
NUMBER OF SEQUENCES: 117
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: CA
COUNTRY: U.S.A.
ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/800,644
FILING DATE: 14-FEB-1997
CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/056,200
FILING DATE: 30-APR-1993

ATTORNEY/AGENT INFORMATION:

NAME: Fedick, Michael P.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 94:
SEQUENCE CHARACTERISTICS:
LENGTH: 1898 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-800-644-94

Query Match 2.3%; Score 244.5; DB 2; Length 1898;
Best Local Similarity 26.8%; Pred. No. 1.2e-09;
Matches 83; Conservative 64; Mismatches 96; Indels 67; Gaps 11;

QY 635 CTKPNQKRPDQDAVNLQRLYSGLMLETVAI-RKAGIAYRRFPDFFKRYVLMKNTA 693
: : : : :
Db 196 CQGHETEPDDE-----EQLRRRLRLRRKRGEEKOCQORERD-----RVFO----- 239
: : : : :
QY 694 LPEDVNGKSTLSQLVYDASNSEWQIGTKYVFLRESLEQKLEKRESEVSHAMVIRAHVL 753
: : : : :
Db 240 -----ESEEKMKRKETVLRKEEKLQEEFPQKRELQEEBQRLK----- 280
: : : : :
QY 754 GFLARKQYKVLVYCVIIQKNTAFLLRRFLHLKKAIVFOQLRGQIARVYRQLAE 813
: : : : :
Db 281 --LEROELRRRQ-----EEOOQOQLRR-----EQQLRK-QEERREQOE 320
: : : : :
QY 814 KREOEKKKQOESEK--KREEF-----REREREALRAQOEETRKQOELTA 863
: : : : :
Db 321 KREOEKKKQOESEK--KREEF-----REREREALRAQOEETRKQOELTA 863

QY 864 LKQKQKAEITRELEKQENK-QVEEILRLKEIEDLQMKQOEELSTEAISLOKQERR 922
: : : : :
Db 381 QEERREQQLRREQQLRREQQLRREQQLRREQQLRREQQLRREQQLR 440
: : : : :
QY 923 DQELRLREE 932
: : : : :
Db 441 EQLRREOE 450

RESULT 14

US-08-910-925-1
; Sequence 1, Application US/08910925
; Patent No. 6162601

GENERAL INFORMATION:

APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN PININ SPLICE VARIANT
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,925
FILING DATE: Herewith
CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0365 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TEXT:

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 717 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: FIBRN0T01
CLONE: 53219
US-08-910-925-1

Query Match 2.0%; Score 218; DB 4; Length 717;
Best Local Similarity 23.0%; Pred. No. 2.9e-08;
Matches 125; Conservative 75; Mismatches 175; Indels 168; Gaps 23;

QY 729 LEOKLE-KRRESEVSHAMVIRAHVGLFARKQYRKVL-----CVIIQKRY 775
: : : : :
Db 209 LEOKVELAQLOEWEHNAKTIK-----YIRTKYPHLYIPGRNCFATQKLIBSQRK 263
: : : : :
QY 776 RAFLLRRLFLHLKKAIVFOQLRGQIARVYRQLAE-----RQSEKQOESEK 830
: : : : :
Db 264 NALFEGRR-----IEFADQI-NKMDARRRQSMKEKHQVYVNEQKAEQEGKVAQ 314
: : : : :
QY 831 REEERERERERREALRAQOEETR-----KQOELALQKQKAEAL-----TR 875
: : : : :
Db 315 REELEEETGNQNDVIEEAGQEESEKKAIVHSDAEKQOE-EEOQKQEWVMEETVEVR 373

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:07:56 / Search time 34.8872 Seconds
(without alignments)
4441.378 Million cell updates/sec

Title: US-09-815-379-6

Perfect score: 3898

Sequence: 1 MRLGAAVAALVGRGRAPAS.....INKYLEANGQLPVKKKAKN 752

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTRMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriopl:*
- 17: sp_archaeop:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3841.5	98.6	751	4 Q96T39	Q96T39 homo sapien
2	3833.5	98.3	751	4 Q96RP9	Q96RP9 homo sapien
3	3506.5	90.0	751	11 Q921D6	Q921D6 mus musculus
4	3499.5	89.8	751	11 Q921D6	Q921D6 mus musculus
5	2521.5	64.7	729	5 Q9VW33	Q9VW33 dirosophila
6	2119.5	54.4	750	5 Q9XV52	Q9XV52 caenorhabdit
7	1990	51.1	757	10 Q941E1	Q941E1 oryza sativ
8	1975.5	50.7	770	10 Q941E1	Q941E1 oryza sativ
9	1972	50.6	754	10 Q9C641	Q9C641 arabidopsis
10	1970	50.5	754	10 Q9SHD6	Q9SHD6 arabidopsis
11	1964	50.4	757	10 Q9FEG4	Q9FEG4 oryza sativ
12	1947.5	50.0	757	3 Q96WU0	Q96WU0 arxula aden
13	1472	37.8	692	16 Q8YP62	Q8YP62 arabidopsis
14	1470	37.7	682	2 Q9RNG0	Q9RNG0 heliobacter
15	1427.5	36.6	700	2 Q9AF78	Q9AF78 archibacter
16	1424.5	36.5	693	16 Q8R602	Q8R602 fusobacteri

17	1422.5	36.5	690	16 Q8R7V1	Q8R7V1 thermosae
18	1416.5	36.3	694	2 Q8RTE9	Q8RTE9 bartonella
19	1407	36.1	701	2 Q8RME2	Q8RME2 mycobacteri
20	1398	35.9	699	16 Q8UE15	Q8UE15 agrobacteri
21	1372.5	35.2	692	2 Q9F4B2	Q9F4B2 bacillus st
22	1367	35.1	695	16 Q9KPM5	Q9KPM5 vibrio chol
23	1365	35.0	707	2 Q9RHH6	Q9RHH6 porphyromon
24	1360	34.9	707	2 Q9RHH7	Q9RHH7 porphyromon
25	1358	34.8	707	2 Q9RHH6	Q9RHH6 porphyromon
26	1356	34.8	707	2 Q9RHH4	Q9RHH4 porphyromon
27	1354	34.7	707	2 Q9RHH5	Q9RHH5 porphyromon
28	1351.5	34.7	558	3 Q9VUA3	Q9VUA3 schizosacch
29	1341.5	34.4	783	10 Q9S175	Q9S175 arabidopsis
30	1337.5	34.3	689	2 Q9VMU0	Q9VMU0 mycoplasma
31	1334.5	34.2	783	10 Q94BR7	Q94BR7 arabidopsis
32	1142.5	29.3	681	2 Q9A1G7	Q9A1G7 carsonella
33	1074.5	27.6	696	16 Q8XP86	Q8XP86 clostridium
34	1074	27.6	700	16 Q8R7R5	Q8R7R5 thermosae
35	1015.5	26.1	779	4 Q969S9	Q969S9 homo sapien
36	988	25.3	777	11 Q8R2Q4	Q8R2Q4 mus musculu
37	959	24.6	732	4 Q8WY10	Q8WY10 homo sapien
38	921.5	23.6	690	16 Q8R604	Q8R604 fusobacteri
39	803.5	20.6	678	16 Q9RXC2	Q9RXC2 deinococcus
40	786.5	20.2	813	3 Q94429	Q94429 schizosacch
41	784.5	20.1	1093	5 Q9VCX4	Q9VCX4 dirosophila
42	754	19.3	689	5 Q95Y73	Q95Y73 caenorhabdi
43	737	18.9	694	16 Q8YPI9	Q8YPI9 arabidopsis
44	709	18.2	513	4 Q9H6Z1	Q9H6Z1 homo sapien
45	700	18.0	236	5 Q00887	Q00887 leishmania

ALIGNMENTS

RESULT 1	ID	Q96T39	PRELIMINARY	PRT	751 AA
AC	Q96T39	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)				
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)				
DE	Elongation factor G1.				
EFGL					
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Butelostomi;				
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OK	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	Hammarund M., Wilson W., Corcoran W., Binhorn S., Grander D.,				
RA	Sangfelt O.,				
RT	"Identification and characterization of two novel human elongation				
RT	factor genes, hERG2 and hERG1, phylogenetically conserved through				
RT	evolution."				
RL	Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.				
DR	EMBL; AF367998; AAK53402.1; -				
DR	InterPro; IPR004540; EF-G				
DR	InterPro; IPR000640; EF-G_C				
DR	InterPro; IPR004161; EFU_D2				
DR	InterPro; IPR000795; EF_GTPbind				
DR	InterPro; IPR005225; Small_GTP				
DR	Pfam; PF00679; EFG_G_1				
DR	Pfam; PF00009; GTP_EFTU_1				
DR	Pfam; PF03144; GTP_EFTU_D2_1				
DR	TIGRFAMs; TIGR00484; EF-G_1				
DR	TIGRFAMs; TIGR00231; Small_GTP_1				
DR	PROSITE; PS00301; EFATOR GTP; UNKNOWN_1				
DR	GTP-binding; Protein biosynthesis				
DR	SEQUENCE 751 AA; 83471 MW; 5937FFB2A089E2E CRC64;				
Query Match	98.6%; Score 3841.5; DB 4; Length 751;				
Best Local Similarity	98.8%; Pred. No. 2.3e-258;				
Matches 743; Conservative	2; Mismatches 6; Indels 1; Gaps 1;				


```

QY 1 MRLGAAVAALGRGRAPASLGMQRKQVNMWKA CRSSSGVTPNEKIRNIGISAHIDSGKT 60
DB 1 MRLGAAVAALGRGRAPASLGMQRKQVNMWKA CRSSSGVTPNEKIRNIGISAHIDSGKT 60
QY 61 TLTERVLYYTGRIAKHNEVKGKDGAVAMDSELEBRGITTQSAATFTMMKDVNINIID 120
DB 61 TLTERVLYYTGRIAKHNEVKGKDGAVAMDSELEBRGITTQSAATFTMMKDVNINIID 120
QY 121 TPGHVDFTIEVERALRVLDAVLVCAVGVQCQTMTVNRQMRVNVPLFTINKLDRMG 180
DB 121 TPGHVDFTIEVERALRVLDAVLVCAVGVQCQTMTVNRQMRVNVPLFTINKLDRMG 180
QY 181 SNPARALQOMRSLKLNHNAFMQIPMGLEGNFKIYDLIBERAIYFDGDSQIVRYGEIPA 240
DB 181 SNPARALQOMRSLKLNHNAFMQIPMGLEGNFKIYDLIBERAIYFDGDSQIVRYGEIPA 240
QY 241 ELRAAATDHRQELIECVANSDEQLGEMPLEEKIPIISDLKLAIRATLKRSFTPVFLGSA 300
DB 241 ELRAAATDHRQELIECVANSDEQLGEMPLEEKIPIISDLKLAIRATLKRSFTPVFLGSA 300
QY 301 LKNKGVOPLDVALELVPNSEVQNYAAILNKDSDSEKTKILMNSSRNSHPVGLAPFL 360
DB 301 LKNKGVOPLDVALELVPNSEVQNYAAILNKDSDSEKTKILMNSSRNSHPVGLAPFL 360
QY 361 EVGRFGQLTYVRSYOGELKKGDTTYNTRTKKVRLORLARHADMMEASTEEVYAGDICA 420
DB 361 EVGRFGQLTYVRSYOGELKKGDTTYNTRTKKVRLORLARHADMMEASTEEVYAGDICA 420
QY 421 LFGIDCAGDFTTDKANSGLSMESIHVPDPVVISIMKPSNKNDLEKFSKIGRFTREDPT 480
DB 421 LFGIDCAGDFTTDKANSGLSMESIHVPDPVVISIMKPSNKNDLEKFSKIGRFTREDPT 480
QY 481 FKYPFDTENKETVISGMGELHLEIYAQRLERBYGCPCTIGKPKVAFRETTAPVPDFTH 540
DB 481 FKYPFDTENKETVISGMGELHLEIYAQRLERBYGCPCTIGKPKVAFRETTAPVPDFTH 540
QY 541 KKQSGAGQYGVKIVLEPLDPEDYTKLEFSDETFSGSI PKQFPAVEKGFADACEKGL 600
DB 541 KKQSGAGQYGVKIVLEPLDPEDYTKLEFSDETFSGSI PKQFPAVEKGFADACEKGL 600
QY 601 SGHKSGLRFLVODGAHHWDSNEISFIRAGBALKQALANATLCLIEPIMAVEVVAANE 660
DB 601 SGHKSGLRFLVODGAHHWDSNEISFIRAGBALKQALANATLCLIEPIMAVEVVAANE 660
QY 661 FQGVYAGINRRHGVITGQDGVEDYFTLYADVPVLMDFGYSLEBSCTEGKGEYTMESYR 720
DB 661 FQGVYAGINRRHGVITGQDGVEDYFTLYADVPVLMDFGYSLEBSCTEGKGEYTMESYR 720
QY 721 YOPCLPSTQEDVINKYLEATGQLPVKKGAKN 752
DB 721 YOPCLPSTQEDVINKYLEATGQLPVKKGAKN 752

```

```

RESULT 2
Q96RP9 PRELIMINARY; PRT; 751 AA.
AC 096RP9;
DT 01-DEC-2001 (TEMBLrel. 19, Created)
DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DB 01-DUN-2002 (TEMBLrel. 21, Last annotation update)
DB Elongation factor G.
GN EFG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21269185; PubMed=11374907;
RT Cloning and Characterization of Human and Mouse Mitochondrial
RT Elongation Factor G, Gfm and Gfm, and Mapping of Gfm to Human

```

```

RT Chromosome 3q25.1-q26.2.;
RL Genomics 74:109-114(2001).
DR EMBL; AF309777; AAK58877.1; -.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EFG-C.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00679; EFG_C_1.
DR Pfam; PF00009; GTP_EFTU_1.
DR Pfam; PF03144; GTP_EFTU_D2_1.
DR TIGRFAMs; TIGR00484; EF-G_1.
DR TIGRFAMs; TIGR00231; small_GTP_1.
DR PROSITE; PS00301; EFATOR GTP; UNKNOWN_1.
KW GTP-binding; Protein biosynthesis.
SQ SEQUENCE 751 AA; 83505 MW; BF794C925A6775C2 CRC64;

```

```

Query Match 98.3%; Score 3833.5; DB 4; Length 751;
Best Local Similarity 98.5%; Pred. No. 8.2e-258;
Matches 741; Conservative 3; Mismatches 7; Indels 1; Gaps 1;

```

```

QY 1 MRLGAAVAALGRGRAPASLGMQRKQVNMWKA CRSSSGVTPNEKIRNIGISAHIDSGKT 60
DB 1 MRLGAAVAALGRGRAPASLGMQRKQVNMWKA CRSSSGVTPNEKIRNIGISAHIDSGKT 60
QY 61 TLTERVLYYTGRIAKHNEVKGKDGAVAMDSELEBRGITTQSAATFTMMKDVNINIID 120
DB 61 TLTERVLYYTGRIAKHNEVKGKDGAVAMDSELEBRGITTQSAATFTMMKDVNINIID 120
QY 121 TPGHVDFTIEVERALRVLDAVLVCAVGVQCQTMTVNRQMRVNVPLFTINKLDRMG 180
DB 121 TPGHVDFTIEVERALRVLDAVLVCAVGVQCQTMTVNRQMRVNVPLFTINKLDRMG 180
QY 181 SNPARALQOMRSLKLNHNAFMQIPMGLEGNFKIYDLIBERAIYFDGDSQIVRYGEIPA 240
DB 181 SNPARALQOMRSLKLNHNAFMQIPMGLEGNFKIYDLIBERAIYFDGDSQIVRYGEIPA 240
QY 241 ELRAAATDHRQELIECVANSDEQLGEMPLEEKIPIISDLKLAIRATLKRSFTPVFLGSA 300
DB 241 ELRAAATDHRQELIECVANSDEQLGEMPLEEKIPIISDLKLAIRATLKRSFTPVFLGSA 300
QY 301 LKNKGVOPLDVALELVPNSEVQNYAAILNKDSDSEKTKILMNSSRNSHPVGLAPFL 360
DB 301 LKNKGVOPLDVALELVPNSEVQNYAAILNKDSDSEKTKILMNSSRNSHPVGLAPFL 360
QY 361 EVGRFGQLTYVRSYOGELKKGDTTYNTRTKKVRLORLARHADMMEASTEEVYAGDICA 420
DB 361 EVGRFGQLTYVRSYOGELKKGDTTYNTRTKKVRLORLARHADMMEASTEEVYAGDICA 420
QY 421 LFGIDCAGDFTTDKANSGLSMESIHVPDPVVISIMKPSNKNDLEKFSKIGRFTREDPT 480
DB 421 LFGIDCAGDFTTDKANSGLSMESIHVPDPVVISIMKPSNKNDLEKFSKIGRFTREDPT 480
QY 481 FKYPFDTENKETVISGMGELHLEIYAQRLERBYGCPCTIGKPKVAFRETTAPVPDFTH 540
DB 481 FKYPFDTENKETVISGMGELHLEIYAQRLERBYGCPCTIGKPKVAFRETTAPVPDFTH 540
QY 541 KKQSGAGQYGVKIVLEPLDPEDYTKLEFSDETFSGSI PKQFPAVEKGFADACEKGL 600
DB 541 KKQSGAGQYGVKIVLEPLDPEDYTKLEFSDETFSGSI PKQFPAVEKGFADACEKGL 600
QY 601 SGHKSGLRFLVODGAHHWDSNEISFIRAGBALKQALANATLCLIEPIMAVEVVAANE 660
DB 601 SGHKSGLRFLVODGAHHWDSNEISFIRAGBALKQALANATLCLIEPIMAVEVVAANE 660
QY 661 FQGVYAGINRRHGVITGQDGVEDYFTLYADVPVLMDFGYSLEBSCTEGKGEYTMESYR 720
DB 661 FQGVYAGINRRHGVITGQDGVEDYFTLYADVPVLMDFGYSLEBSCTEGKGEYTMESYR 720
QY 721 YOPCLPSTQEDVINKYLEATGQLPVKKGAKN 752
DB 721 YOPCLPSTQEDVINKYLEATGQLPVKKGAKN 752

```

RESULT 3
ID 092106 PRELIMINARY; PRT: 751 AA.
AC 092106:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN Unknown (protein for MG:7961).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Straube R. J.
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013093; AAH13093.1; -.
DR MGD; MG1:107339; GfM.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EF-G.
DR InterPro; IPR004161; EFTU D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00679; EFG_C_1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU D2; 1.
DR TIGRFAMs; TIGR00484; EF-G; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFACITOR_GTP; UNKNOWN_1.
KW GTP-binding; Protein biosynthesis.
SQ SEQUENCE 751 AA; 83539 MW; 1DCE4D8981EAD2C CRC64;

Query Match 90.0%; Score 3506.5; DB 11; Length 751;
Beet Local Similarity 89.4%; Pred. No. 4.4e-235;
Matches 675; Conservative 34; Mismatches 39; Indels 7; Gaps 4;

1 MRLGAAVAALGRG---RAPASLGQKQVNMKACRWSSSGVIPNEKIRNIGISAHIDS 57
1 MRLGAAVAALGRG---RAPASLGQKQVNMKACRWSSSGVIPNEKIRNIGISAHIDS 57
1 MRLGAAVAALGRG---RAPASLGQKQVNMKACRWSSSGVIPNEKIRNIGISAHIDS 57
58 GKTLLTERLVYTGRIAGMEVKGKDGAVGAVMDSMELERQGITTSAAATFTMKDVIN 117
59 GKTLLTERLVYTGRIAGMEVKGKDGAVGAVMDSMELERQGITTSAAATFTMKDVIN 118
118 IIDTGHVDFTEVERALRVLDGAVLVCAVGVQCCQMTVNRQKRYNVPLTFTINKD 177
119 IIDTGHVDFTEVERALRVLDGAVLVCAVGVQCCQMTVNRQKRYNVPLTFTINKD 178
178 RMGSPARALQOMRSKLNHNTAFMQIPMGLGEGNFKGIVDLIERAIYFDGDSQIVRYGE 237
179 RMGSPARALQOMRSKLNHNTAFMQIPMGLGEGNFKGIVDLIERAIYFDGDSQIVRYGE 238
238 IPAEIRAAATDHRQELIECVANSDEQLGEMPLEEKIPSIDKLAIIRATLKRSFTPYFL 297
239 IPAGRAAAADHRQELIECVANSDEQLGEMPLEEKIPSIDKLAIIRATLKRSFTPYFL 298
298 GSALNKQGVQPLLDVAVLEVLPNPSEVQNTAIIANKKDSSEKTKILMNSRRHNSHPVGIA 357
299 GSALNKQGVQPLLDVAVLEVLPNPSEVQNTAIIANKKDSSEKTKILMNSRRHNSHPVGIA 357
358 PFLVGRGQQLTVYVSYQELKKKDTIYTRTKRYRLRLARMHADMMEAESTEVEYVGD 417
359 PFLVGRGQQLTVYVSYQELKKKDTIYTRTKRYRLRLARMHADMMEAESTEVEYVGD 417
358 PFLVGRGQQLTVYVSYQELKKKDTIYTRTKRYRLRLARMHADMMEAESTEVEYVGD 417
359 PFLVGRGQQLTVYVSYQELKKKDTIYTRTKRYRLRLARMHADMMEAESTEVEYVGD 417
418 ICALFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSNKDLEKFSKIGFTRE 477
419 ICALFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSNKDLEKFSKIGFTRE 477
417 ICALFGIDCASGDTFTDKANSGLSMESIHVPDVISIANKPSNKDLEKFSKIGFTRE 476
478 DPTFKVYPTENKEVYISGMGELHLEIYQRLEREGVCTITGKPKVARETITAPVPD 537
479 DPTFKVYPTENKEVYISGMGELHLEIYQRLEREGVCTITGKPKVARETITAPVPD 537
477 DPTFKVYPTENKEVYISGMGELHLEIYQRLEREGVCTITGKPKVARETITAPVPD 536

538 FTHKKQSGAGQYGVIVLEPLDPEDYTKLEFSDTEFGSNI PKQFPAVEKGFIDACEK 597
539 FTHKKQSGAGQYGVIVLEPLDPEDYTKLEFSDTEFGSNI PKQFPAVEKGFIDACEK 596
598 GPLSGHKLSGLRFLVQDGAHNMVDSNEISFTRAGEGALKOLANATLCLIEPIANVEYVA 657
599 GPLSGHKLSGLRFLVQDGAHNMVDSNEISFTRAGEGALKOLANATLCLIEPIANVEYVA 656
658 PNEFGQVYIAGINRHRGVTGDDVEDYFTLYADVPVLDMMRGYSTELASCTEGKGEYTM 717
659 PNEFGQVYIAGINRHRGVTGDDVEDYFTLYADVPVLDMMRGYSTELASCTEGKGEYTM 716
718 YSRVQCLPSTQEDVYINKYLEATGQLPYKKGAKN 752
719 YSRVQCLPSTQEDVYINKYLEATGQLPYKKGAKN 751
717 YSRVQCLPSTQEDVYINKYLEATGQLPYKKGAKN 751

RESULT 4
ID 092410 PRELIMINARY; PRT: 751 AA.
AC 092410:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
GN Elongation factor G.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21269185; PubMed=11374907;
RA Gao J., Yu L., Zhang P., Jiang J., Chen J., Peng J., Wei Y., Zhao S.;
RT "Cloning and Characterization of Human and Mouse Mitochondrial
RT Elongation Factor G, GfM and GfM and Mapping of GfM to Human
RT Chromosome 3q25.1-q26.2.";
RT Genomics 74:109-114(2001).
DR EMBL; AF15511; AAKS8878.1; -.
DR MGD; MG1:107339; GfM.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EF-G.
DR InterPro; IPR004161; EFTU D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00679; EFG_C_1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU D2; 1.
DR TIGRFAMs; TIGR00484; EF-G; 1.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFACITOR_GTP; UNKNOWN_1.
KW GTP-binding; Protein biosynthesis.
SQ SEQUENCE 751 AA; 83667 MW; C206377C48E97808 CRC64;

Query Match 89.8%; Score 3499.5; DB 11; Length 751;
Beet Local Similarity 89.1%; Pred. No. 1.4e-234;
Matches 673; Conservative 36; Mismatches 39; Indels 7; Gaps 4;

1 MRLGAAVAALGRG---RAPASLGQKQVNMKACRWSSSGVIPNEKIRNIGISAHIDS 57
1 MRLGAAVAALGRG---RAPASLGQKQVNMKACRWSSSGVIPNEKIRNIGISAHIDS 57
1 MRLGAAVAALGRG---RAPASLGQKQVNMKACRWSSSGVIPNEKIRNIGISAHIDS 58
58 GKTLLTERLVYTGRIAGMEVKGKDGAVGAVMDSMELERQGITTSAAATFTMKDVIN 117
59 GKTLLTERLVYTGRIAGMEVKGKDGAVGAVMDSMELERQGITTSAAATFTMKDVIN 118
118 IIDTGHVDFTEVERALRVLDGAVLVCAVGVQCCQMTVNRQKRYNVPLTFTINKD 177
119 IIDTGHVDFTEVERALRVLDGAVLVCAVGVQCCQMTVNRQKRYNVPLTFTINKD 178
178 RMGSPARALQOMRSKLNHNTAFMQIPMGLGEGNFKGIVDLIERAIYFDGDSQIVRYGE 237

Db 179 RMGNSBRALQOMRSKLNHNAAFVQIPIGLEDFKGIIDLIBERRAIFPDGFCQIVRYDB 238
 Qy 238 IPAELRAAATDHROCELECVANSDEQGEFLEBKIPISIDLTALIRATLKASFTPVFL 297
 Db 239 IIPGLAAMAAADHROCELECVANSDEQGEFLEBKIPISIDLTALIRATLKASFTPVFL 298
 Qy 298 GSALKNKKGVQPLLDVALEYLNPSEVONYAIIINKKDSKEKTKILMNSRHNSHPVGLA 357
 Db 299 GSALKNKKGVQPLLDVALEYLNPSEVONYAIIINKKDSKEKTKILMNSRHNSHPVGLA 357
 Qy 358 FPLEVBRFQGLTYVRSYQGLKKGDTTYNTRTKKRYRLQRLAMADHMAEASIEEYVAD 417
 Db 358 FPLEVBRFQGLTYVRSYQGLKKGDTTYNTRTKKRYRLQRLAMADHMAEASIEEYVAD 416
 Qy 418 ICALFIDCASGDTFIDKANSGLSMESIHVPDPVSIAMKPSKNDLEKFSKIGFTBE 477
 Db 417 ICALFIDCASGDTFIDKANSGLSMESIHVPDPVSIAMKPSKNDLEKFSKIGFTBE 476
 Qy 478 DPTFKVYPTENKTEVVISGMLHLIYVQRLEREXGCPCTITGKPRVARETITAPVPP 537
 Db 477 DPTFKVHPDESKETIVSGMGLHLIYVQRLEREXGCPCTITGKPRVARETITAPVPP 536
 Qy 538 FTHKKSQSGAGQVGVKIVLEPLDPEDDYTKLESDTFPSNIPKQVPAVEKGFLLDACX 597
 Db 537 FTHKKSQSGAGQVGVKIVLEPLDPEDDYTKLESDTFPSNIPKQVPAVEKGFLLDACX 596
 Qy 598 GPLSGHKLGLRFLVLDGAMHWDSNEISFIRAGEGLKQALNATLCTILEPIMAVEYVA 657
 Db 597 GPLSGHKLGLRFLVLDGAMHWDSNEISFIRAGEGLKQALNATLCTILEPIMAVEYVA 656
 Qy 658 PNEFGQVVIAGINRRHGVITGDQGVEDYFTLVADVLNDFGYSTELRSCTECKGEYME 717
 Db 657 PNEFGQVVIAGINRRHGVITGDQGVEDYFTLVADVLNDFGYSTELRSCTECKGEYME 716
 Qy 718 YSRVOPCLSTQEDVINKYLEATGOLPYKKGKAKN 752
 Db 717 YSRVOPCLSTQEDVINKYLEATGOLPYKKGKAKN 751
 RESULT 5
 Qy 09VM33 PRELIMINARY; PRT; 729 AA.
 AC 09VM33;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE CG4567 protein.
 GN CG4567.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN 11)
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abail J.F., Achapant A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
 RA Ballew R.M., Beasly P., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Betman B.P., Bhandari D., Bolshakov S.,
 RA Borke D., Bouchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.B., Gary N.S., Gelbart W.M., Glasser K.,

RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jajuli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Krafc C., Kravitz S., Kulp D., Lai Z.,
 RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mactel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusse K.D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reiter K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spieding A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan W., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of Drosophila melanogaster";
 Science 287:2185-2195 (2000).
 DR EMBL: AE003616; AAF52495.1; -.
 DR HSSP: P1351; IFNM.
 DR Flybase; FBgn0031898; CG4567.
 DR InterPro; IPR004540; EF-G.
 DR InterPro; IPR000640; EFG-C.
 DR InterPro; IPR004161; EFTU D2.
 DR InterPro; IPR000785; EF_GTPbind.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF00679; EFG_C_1.
 DR Pfam; PF00009; GTP_BPTU_1.
 DR Pfam; PF03144; GTP_BPTU_D2_1.
 DR PRINTS; PR00315; E1ONGATNFC.
 DR TIGRFAMs; TIGR00484; EF-G_1.
 DR TIGRFAMs; TIGR00231; small_GTP_1.
 DR TIGRFAMs; TIGR00231; small_GTP_1.
 DR PROSITE; PS00301; EFACITOR_GTP_1.
 DR GTP-binding; Protein biosynthesis.
 SQ SEQUENCE 729 AA; 81739 MW; 6CB15A76DF1807AB CRC64;
 Query Match 64.7%; Score 2521.5; DB 5; Length 729;
 Best Local Similarity 66.1%; Pred. No. 1.2e-166;
 Matches 469; Conservative 101; Mismatches 136; Indels 3; Gaps 2;
 Qy 44 EKIRNGISAHIDSCKTLTERLVLYYGRIRAHKEVKGQGVAMDSMELEKRGITIQ 103
 Db 24 KIRNIGISAHIDSCKTLTERLVLYYGRIRAHKEVKGQGVAMDSMELEKRGITIQ 83
 Qy 104 SAATFTMKVDVINIIDTPGHVDFTEVERALRVLDGAVLVCAVGVQCQTMVNRQMK 163
 Db 84 SAATFTMKVDVINIIDTPGHVDFTEVERALRVLDGAVLVCAVGVQCQTMVNRQMK 143
 Qy 164 RYVNFPLFINKLDMKSNIPARALQOMRSKLNHNAFQIIPMGLEGNFRGIYDLLEBAI 223
 Db 144 RYVNFPLFINKLDMKSNIPARALQOMRSKLNHNAFQIIPMGLEGNFRGIYDLLEBAI 203
 Qy 224 YVDGFSQIVRGEIPAEELRAAATDHROCELECVANSDEQGEFLEBKIPISIDLTAL 283
 Db 204 YVEGHGMDIRDELIPDMKRVSLERQELIELHLSNABETLGELEBKEPREDIDIKAL 263
 Qy 284 RRATLKRSFTPVFLGASALKNKKGVQPLLDVALEYLNPSEVONYAIIINKKDSKEKTKILM 343
 Db 264 RRATLKRSFTPVFLGASALKNKKGVQPLLDVALEYLNPSEVONYAIIINKKDSKEKTKILM 321
 Qy 344 NSSRNSHPFVGLAPPLVGRGQGLTYVRSYQGLKKGDTTYNTRTKKRYRLQRLARMA 403
 Db 322 NPARGKQPFVGLAPPLVGRGQGLTYVRSYQGLKKGDTTYNTRTKKRYRLQRLARMA 381
 Qy 404 DWKEASTEEVYAGDICALFGIDCASGDTFIDKANSGLSMESIHVPDPVSIAMKPSKND 463
 Db 382 NQME-DVNEVYAGDICALFGIDCASGDTFIDKANSGLSMESIHVPDPVSIAMKPSKND 440
 Qy 464 LEKFSKIGRFTREDPTFKVYPTENKTEVVISGMLHLIYVQRLEREXGCPCTITGKPR 523

DR InterPro: IPR005225; Small_GTP:
DR Pfam: PF00679; EFG_C, 1.
DR Pfam: PF00009; GTP_EFTU, 1.
DR Pfam: PF03144; GTP_EFTU_D2, 1.
DR TIGRfam: TIGR00464; EF-G, 1.
DR TIGRfam: TIGR00231; small_GTP, 1.
DR PROSITE: PS00301; EFATOR_GTP; UNKNOWN 1.
KW Elongation factor; GTP-binding; Protein biosynthesis.
KW Elongation factor; GTP-binding; Protein biosynthesis.
SQ SEQUENCE 757 AA; 83924 MW; C3E7C7068ACD7B93 CRC64;

Query Match 51.1%; Score 1990; DB 10; Length 757;
Best Local Similarity 53.6%; Pred. No. 1,2e-129;
Matches 396; Conservative 120; Mismatches 209; Indels 14; Gaps 6;

3 LIGAAVAALGRAPASLGMORQVNMKACRWSSGVIPNEKINIGISAHIDSGKTL 62
1 LSAAAAAAAAARRGMSASA--LRARDEKEVARRES---MDMRNIGISAHIDSGKTL 84
63 TERVLYTGTGRIAMHEVKGKGVAGVAVDSMELEORGITIOSAATFTMMKVNIIDTP 122
85 TERVLYTGTGRIAMHEVKGKGVAGVAVDSMELEORGITIOSAATFTMMKVNIIDTP 144
123 GHVDFTEVERALRVLDGAVLVLCVAVGVOCTMTVNRQMKRYNVPFLTINKLDRMGSN 182
145 GHVDFTEVERALRVLDGAVLVLCVAVGVOCTMTVNRQMKRYNVPFLTINKLDRMGSN 204
183 PARALQOMSKLNHTAFMQIPMGLEGNFNGIYDLIEERAIYFDGDFSGIVRGEPAPL 242
205 PMKVLNQASKLHNNAAVQVPIGLEBEPEGVLVDVLEKAYKEGSGQNVASDVPSNM 264
243 RAAATDHROELIECVANSDEQLGEMPLEEKIPIISDLKAIIRATLKRSFTPVFLGSAIK 302
265 ODLYMKREELIEVSEVDQLAEAFLENDPEIQANQLKAIIRATVARKFIPVYMSARK 324
303 NKGVOLDAVLEVLEPSEVQNYAILNKDDSEKTKILMSRNNSHPVGLAPLEV 362
325 NKGVOLDAVLEVLEPSEVQNYAILNKDDSEKTKILMSRNNSHPVGLAPLEV 378
363 GREGLTYVRSYOGELKKGDTIYNTRTKKVRLQRLAHMAHMADEASTEVEVAGDICALF 422
379 GRGGLTYLRIVGVRKGDPIYNNATGKKIKPRLVRHNSME-DIQENAHGOIVANF 437
423 GIDCASGDTFTDANSGLSMESIHVPDVISIMKPSNNKDLKFSKIGRTREDPTK 482
438 GVDCAAGDFTD-GSVYKWTSMNVBEPWMSLAVSDISKSGQFSKALNRFOKEDPTR 496
483 YVFDENKKEVVISGMELEHLEIAORLEBYGPCCTGPKVAFRTITAPVFDTHKK 542
497 VGIDPESGETTISGMELHLDIYVERIRREYKVDKAVKPRVNFRETTIORAEFDLARK 556
543 QSGGAGQYGVKIVGLPELPDEDTKLEFSDPTFGSNIPKOFVPAVEKGLDACCKPLSG 602
557 QSGGAGQYGVKIVGLPELPDEDTKLEFSDPTFGSNIPKOFVPAVEKGLDACCKPLSG 616
603 HKLSGLRFLVLDGAHNMVDSNEISFRAGEGALKQALANATLCILEPMAVEVADNEFO 662
617 HVENIRIRIVTDGASHAVDSSELAFKLASIAFRQCAAPARVILEPVMKVELKPTERO 676
663 GOVIAGINRHGIVITQDGVEDYFTLYADVPLNDMGYSTELASCTEGEGEYMEISRQ 722
677 GTVTDGMRKKGIIIVNDGEGDVTVAACHVPLNMGYSTALSKMTQKGEEFMELEIN 736
723 PCLPSTQEDVINKYLEATG 741
737 TVSGDVQMQVLVNTYKASRG 755

RESULT 8

094151 PRELIMINARY; PRT; 770 AA.

AC 094151;
DT 01-DEC-2001 (T-EMBLrel. 19, Created)
DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)

DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
OS Mitochondrial elongation factor G.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;

RA SEQUENCE FROM N.A.
RA Huang E.N., de la Bastide M., Preston R.R., Nascimento L.U.,
RA Spiegel L.A., Vil M.D., Baker J.P., Shah R.S., Kirchhoff K.A.,
RA Rodriguez M.A., King L., O'Shaughnessy A., Bal H., Dedhia N.N.,
RA McCombie W.R.;
RT "Genomic Sequence For Oryza sativa, Nipponbare strain, Clone
RT OSUB0084C09, Complete Sequence";
RN Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.

[2]

RA SEQUENCE FROM N.A.
RA McCombie W.R.;
RN Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.

[3]

RA Palmer L.E., Bal H., Huang E.N., de la Bastide M., Preston R.R.,
RA Nascimento L.U., Spiegel L.A., Vil M.D., Baker J.P., Shah R.S.,
RA Kirchhoff K.A., Rodriguez M.A., King L., O'Shaughnessy A., Dedhia N.N.,
RA McCombie W.R.;
RN Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

[4]

RA SEQUENCE FROM N.A.
RA Palmer L.E., Bal H., Huang E.N., de la Bastide M., Preston R.R.,
RA Nascimento L.U., Spiegel L.A., Vil M.D., Baker J.P., Shah R.S.,
RA Kirchhoff K.A., Rodriguez M.A., King L., O'Shaughnessy A., Dedhia N.N.,
RA McCombie W.R.;
RN Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL: AC016781; AAK53868.1; --
DR InterPro: IPR004540; EF-G.
DR InterPro: IPR00640; EF-G.
DR InterPro: IPR004161; EFTU_D2.
DR InterPro: IPR00795; EF_GTPbind.
DR InterPro: IPR00525; Small_GTP.
DR Pfam: PF00679; EFG_C, 1.
DR Pfam: PF00009; GTP_EFTU, 1.
DR Pfam: PF03144; GTP_EFTU_D2, 1.
DR TIGRfam: TIGR00484; EF-G, 1.
DR TIGRfam: TIGR00231; small_GTP, 1.
DR PROSITE: PS00301; EFATOR_GTP; UNKNOWN 1.
KW Elongation factor; GTP-binding; Protein biosynthesis.
KW Elongation factor; GTP-binding; Protein biosynthesis.
SQ SEQUENCE 770 AA; 85268 MW; 94D0D95F5119606A CRC64;

Query Match 50.7%; Score 1975.5; DB 10; Length 770;
Best Local Similarity 52.7%; Pred. No. 1,2e-128;
Matches 396; Conservative 121; Mismatches 208; Indels 27; Gaps 7;

3 LIGAAVAALGRAPASLGMORQVNMKACRWSSGVIPNEKINIGISAHIDSGKTL 62
1 LSAAAAAAAAARRGMSASA--LRARDEKEVARRES---MDMRNIGISAHIDSGKTL 84
63 TERVLYTGTGRIAMHEVKGKGVAGVAVDSMELEORGITIOSAATFTMMKVNIIDTP 122
85 TERVLYTGTGRIAMHEVKGKGVAGVAVDSMELEORGITIOSAATFTMMKVNIIDTP 144
123 GHVDFTEVERALRVLDGAVLVLCVAVGVOCTMTVNRQMKRYNVPFLTINKLDRMGSN 182
145 GHVDFTEVERALRVLDGAVLVLCVAVGVOCTMTVNRQMKRYNVPFLTINKLDRMGSN 204
183 PARALQOM-----RSKLNHTAFMQIPMGLEGNFNGIYDLIEERAIYFDGDF 229
205 PMKVLNQASKLHNNAAVQVPIGLEBEPEGVLVDVLEKAYKEGSGQNVASDVPSNM 264
230 SQIVYRGEPAPLRAATDHROELIECVANSDEQLGEMPLEEKIPIISDLKAIIRATLK 289
265 GQNVASDVPSNMQLVMEKRELEIVSEVDQLAEAFLENDPEIQANQLKAIIRATV 324

```

OY 290 RSFTVEFGSALXKXNGVQGLDLAVETLEPNPSPXVUAYALLINKKDSKEKTKILMNSSRHN 349
Db 325 RKFTIYVMGSAKXMGVQPLDGOVDLYLCEPMVESEYAL---DQNKSEKVLIACT--P 378
OY 350 SHPVPGLAVPLEVGFRGQULTYVRSYOGELKKGDTIYNTRIKKVRLORLARMHADNMEAS 409
Db 379 AEPLVALAFKLEEGFRGQULTYLRIVGVIRKRGDFIVNVTGKKIKVPLRLRMHNSNEME-D 437
OY 410 TEEVYAGDICALFGIDCAGSDPTDQANGLSMESIHVPDPVSIAMKPNKNDLEKFSK 465
Db 438 IOEALAGQVAAFGVDCASGDTFTD-GSVKYYTMTSNWVPEPNWISLAVPSISKSGQFSK 466
OY 470 GIGRTREDPTPKVYFDTENKTEVTJSGMELHLLEIAORLEREYGCPCITGKDPKVAFRET 529
Db 497 ALNRQKEDPTFRVGLDPSGERTIISGMELHLDIYERIRRYKXKADAKYGRVAFRET 556
OY 530 ITAPVPFDPETHKKQSGGAGQYQKVGICVLEPLBEDDYTKLFSDETGSGNIPKQFVPAVEK 589
Db 557 ITORAEFPYLLHKKQSGGQGYGVCVIEPLPESDSGKEEFDMMIIGOALPSNFIPIAIEK 616
OY 590 GPLDCBCKPGLSGHKLGLGRFVLODGAHHMVONSLSPIRAGEGALKOLANLANTLCLLEP 649
Db 617 GPKELCNGSGLGHPVENRIRYLITDASHAIVDSSELAFPLASTIYARQCYAAARPVILEP 676
OY 650 IMAVEVAVPNFEQGOVIAGINRRHGVTIGODGVEDYFTIYADVPILNDFMGYSTELRSCTE 709
Db 677 VMKVELKVPTEBQGVITGDDMNKRKGIIVGNDDGGDDTVVCHVPLNNMFGYSTALSMTG 736
OY 710 GKGEYTMESYQPCPLSTQEDVINYKLYEATG 741
Db 737 GKGEBSMELEHNVTVSODVOMOLVNTYKASRG 768

```

Query Match	Score	Length	50.64	Score 1972	DB 10	Length 754
09C641	PRELIMINARY	PRT	754 AA.			
AC 09C641						
AC 09C641						
DT 01-JUN-2001	(TRMBLrel, 17, Created)					
DT 01-JUN-2001	(TRMBLrel, 17, Last sequence update)					
DT 01-JUN-2002	(TRMBLrel, 21, Last annotation update)					
DE Mitochondrial elongation factor, putative.						
GN P2619.2.						
OS Arabidopsis thaliana (Mouse-ear cress).						
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;						
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.						
OX NCBI TaxID=3702;						
RN [1]						
RP SEQUENCE FROM N.A.						
RC STRAIN=cv. COLUMBIA.						
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,						
RA Maltz R., Ronning C.M., Koo H., Fujii C.Y., Utterback T.R.,						
RA Banksread M.B., Bowman C.H., White O., Nieman W.C., Fraser C.M.;						
RT "Arabidopsis thaliana chromosome 1 BAC F2G19 genomic sequence.";						
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.						
DR EMBL; AC083835; AAG50635.1; .						
DR HSSP; P13551; 1ELO.						
DR InterPro; IPR004540; EF-G.						
DR InterPro; IPR000640; EFG.						
DR InterPro; IPR004161; EFTU D2.						
DR InterPro; IPR000795; EF_GTPbind.						
DR InterPro; IPR005225; Small_GTP.						
DR Pfam; PF00679; EFG_C.1.						
DR Pfam; PF00009; GTP_EFTU.1.						
DR Pfam; PF03144; GTP_EFTU D2.1.						
DR PRINTS; PR00315; ELONGACTNCT.						
DR TIGRFAMs; TIGR00484; EF-G.1.						
DR TIGRFAMs; TIGR00231; small_GTP.1.						
DR PROSITE; PS00301; EFACITOR_GTP.1.						
KM Elongation factor; GTP-binding; Protein biosynthese.						
SQ SEQUENCE 754 AA; 83178 MW; 39B48A89E66F377 CRC64;						

	Best Local Similarity	54.8%	Pred. No. 2,1e-118		
	Matches	385;	Conservative	116;	Mismatches 187; Indels 12; Gaps 5
Qy	35	WSSGVIENKIRNIGISAHIDSKTLLTEREVLYYTGRIAMHBYKQGVGVANDSMEL	94		
Db	58	WKES----MDLKRNIGISAHIDSKTLLTEREVLPYTGRIHIEHVEGRGDVGAKKDSIDL	113		
Qy	95	ERQGIITQSAATITMMKDVNNINIIDPGHVDFTIEVERALRVLDGAVLVCANGVQCC	154		
Db	114	EREKGIITQSAATCTWMDYKCNIIIDTPGHVDFTIEVERALRVLDGAVLVCANGVQSQ	173		
Qy	155	TMATNRQKRNVNVPFLPTINFLDMGSPARALOOMSKLHNHTAFAHQIIPMGLSENFQGI	214		
Db	174	STVDROKRRREVPRAVAFINFLDNRGADPMKVLNQAARKLHNHSAVAVPGLSENFQGL	233		
Qy	215	VDLIEBRAIVDGDPSQIVRYGEIIPAELEAATHTHROELIECVANSDBQLGEMPLEEKIP	274		
Db	234	IDLHVAYAFPHSGSENVVAGDIIPADMEGLVAEKRRILITVSEVDVDVLAKEFLNDEPV	293		
Qy	275	SISDLKAIIRATLKRSFTPVFLGSALEKNKGVPLLDVLEYLPNPSFVQVYAILNKXD	334		
Db	294	SASELEAIRATTAQTEVPVPMGSAFAFNKGVQPLDGVASFLESEFNVNVAL----DQ	349		
Qy	335	SKEETKILMNSRRNSHPFVGLAPFLVGRGQILTVRSYOGELKKGDTIYNTRRKRYR	394		
Db	350	NNNEERVLTVGSPDC--PLVALAKLEBGRFGOLTYLKVGEVGIKIGDPIINVTGKRIK	407		
Qy	395	LQRLARHADMMEASTEVEVYAGDICALFGIDCAGDPTDANSGLSMESIHVPDPVISI	454		
Db	408	VPLRVLRHNSNME-DIQEHAHQVAVYFGICASGDTTDD-GSVKTYTMSNAPVPMVSL	465		
Qy	455	AMKPSNKNDLEKFSKGIGRTREDPTFKVYEPDTEKNETVISGMBLEIYAQRLEREYG	514		
Db	466	AVQEVSKDSGGQFSKALNRFQKEDPTFRVGLDPSGQIISGMGBLHDIYVERMRREYK	525		
Qy	515	CPCLTGRPKVAFRETIAPVFPDTHKKQSGAGQYGVYIGVLEPLDREDYTKLEFSQET	574		
Db	526	VDATVGRKRVNFRRETIYQRAEFDYIAHKQSGAGQYGVYIGVLEPLDPSGSEKKEPEEMMI	585		
Qy	575	FGSNIPQPOFVAVNKGFLDACKEKPELSGHKLSGRFLVYODGAAHNVDSNEISPTIAPAGGA	634		
Db	586	VGQAIIPSGFPIAIKKGFEKANSSGLGHPVENRIYLTLDASHAVDSSELAKMAIAYIA	645		
Qy	635	LKQALANATLCLILEPINAVVEVAVNEFQGVYIAGINRRHGVIYTGQDGVEDFTLYADVPL	694		
Db	646	FRLCYTARPILEPVLVLEKLPTEBQGTVAGSINKRKGIIVIGDGGDSVITANVPL	705		
Qy	695	NDMFGYSTELRSCTEBGEYTMESRYOCLPSTQEDVYINKY	736		
Db	706	NNMGYSTLSRSMTOGKEPFMEYKSHSAVSEVQAOCLVNAY	747		
RESULT 10					
95SHD6					
ID	Q9SHD6	PRELIMINARY;	PRT;	754	AA.
AC	Q9SHD6;				
DT	01-MAY-2000 (Tremblrel. 13, Created)				
DT	01-MAY-2000 (Tremblrel. 13, Last sequence update)				
DT	01-JUN-2002 (Tremblrel. 21, Last annotation update)				
DE	Putative mitochondrial translation elongation factor G.				
GN	AT2G45030.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eudicots II; Brassicales; Brassicaceae; Arabidopsids.				
NC	NCBI_taxid=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RX	MEDLINE=20083487; PubMed=10617197;				
RA	Lin X., Kaul S., Rounsley S.D., Shea T. P., Benito M.-I., Town C. D.,				
RA	Fuji1 C. Y., Mason T. M., Bowman C. L., Barnstead M. B., Feldblum T. V.,				
RA	Buell C. R., Ketchum K. A., Lee J. J., Rinning C. M., Ko H., Moffat K. S.,				

RA Cronin L.A., Shen M., Vanaken S.E., Unayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carreira A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nieman W.C., White O., Eissen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RA "Sequence and analysis of chromosome 2 of the plant *Arabidopsis*
 RT thaliana,"
 RL Nature 402:761-768 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC007659; AD32833.1; -.
 DR HSSP; P13551; 1ELO.
 DR InterPro; IPR004540; EF-G.
 DR InterPro; IPR000640; EFG_C.
 DR InterPro; IPR004161; EFTU_D2.
 DR InterPro; IPR000795; EF_GTPbind.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF00679; EFG_C; 1.
 DR Pfam; PF00009; GTP_EFTU; 1.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR PRINTS; PR00315; ELONGATINFC.
 DR TIGRPFAM; TIGR00484; EF-G; 1.
 DR TIGRPFAM; TIGR00231; small_gtp; 1.
 DR PROSITE; PS00301; EFACITOR_GTP; 1.
 DR Elongation factor; GTP-binding; Protein biosynthesis.
 KW Elongation factor; GTP-binding; Protein biosynthesis.
 SQ SEQUENCE 754 AA; 83111 MW; D3A5928C9379FB7 CRC64;

Query Match 50.5%; Score 1970; DB 10; Length 754;
 Best Local Similarity 54.8%; Pred. No. 2,8e-128;
 Matches 385; Conservative 117; Mismatches 188; Indels 12; Gaps 5;

35 MSSSGVLPNEKINIGISAHIDSGKTLTFRVLVYGRYAKMEHVKGQGVAMDSMEL 94
 58 WKSS---MDKLNIGISAHIDSGKTLTFRVLVYGRYAKMEHVKGQGVAMDSMEL 113
 95 ERQGITIOSAATFTMKVNINIDTPGVHDTIEVERALRYLDGAVLVCAVGVQCO 154
 114 ERKGITIOSAATFTMKVNINIDTPGVHDTIEVERALRYLDGAVLVCAVGVQCO 173
 155 TMTVNRQMKRYNVPFLTFINKLDRMGSNPARALQMRSKLNTAFTMOIPMGIEGNFKGI 214
 174 SITVDQMRKYEVPRVAFINKLDRMGADPWKYNQARAKLRHSAVVOPIGIEENFOGI 233
 215 VDLIERATYFDDDFQGIYVGEIIPALRYAATDHOELIECVANSDEOGEMLBEKIP 274
 234 IDLIHVAIFTFHSSGSENVVAGDIPADMGLVDKRELIEIVSEVDDVLAEXFLNDEPV 293
 275 SISDLKLAIRATLKRSFTFVPLGSAIKNGVOPLLDAVLEYLPNPSEVONVAILNKD 334
 294 SAELEBARIRATIAKFPVVFVFGSAFKKNGVOPLLDGVVFLPSPNEVNVNVA---DQ 349
 335 SEKKTILMNSSHHNSHPVGLAFPLEVGFQGLTVYRSYQGLKKGDTYNTRTKKCR 394
 350 NNNEEVTTLGSPDG--PLVALAFKLEBGRFGQLTYLRVYEGVTKGDPFINNNTKRIK 407
 395 LQRLAHMADMASTEVEVYAGDICALFGIDCAGGFTPTDKANSGLSMESIHVPDVISI 454
 408 VPLVLRMHSNDM-DIOEHAIGOIVAFGLECAGSGDTFD-GSVKTYTSMANPEEVMSTI 465
 455 AMKPSNKNDLEKESKIGRFTREDPTFKYVFTDENKEVTVISGNGELHLEIYAQRLREXG 514
 466 AVQPVSKDSGGQFSKALNRFQKEDPTFRVGLDPESSGTTISGNGELHLDIYVERMKREX 525
 515 CPBITGKPKVAFETITAPVPDPFTHKQSGGAGQYKIVLEPDLDPEDYTLAESDET 574
 526 VDATVKPRVNFREITTORAEFDYLRHKQSGGAGQYGRVYGYVEPLPPGSKKEKFEENMI 585
 575 FGSNIPKQFVPAVEKGFADCEKGPLSGHLSGLRFLDQGAHMDVNSISFIRAGEG 634
 586 VGOALPSGLFPAIEKGFKAANSGLIGHFVENLRIVLTDGASHAVDSSELAKRMALTYA 645

OY 635 LKQALANATLCLIEPIMAVEVAPNEFOGOVIAGINRHGVTGDDGVEDYFTLYADVPL 694
 DB 646 FRLCYTAARPVILEBVMVELVKPTEFOGTAVGIDINKSKGLIVNGDGDSDSVITANVPL 705
 OY 695 NMFEGYSTELRCECTGKGEYTEMYSRXPCLPSTBEDVINYK 736
 DB 706 NMFEGYSTLRSMTOGKEFTMEYKESAVNSNEVQALVNAY 747

RESULT 11

OYFE64 PRELIMINARY; PRT; 757 AA.

AC OYFE64;
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Mitochondrial elongation factor G.
 GN ME-F-G.
 OS Oryza sativa (japonica cultivar-group).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_Taxid=39947;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. NIPPONBARE;
 RA Kato A., Fujita S., Komeda Y.;
 RT "Identification and characterization of the gene encoding the
 RT mitochondrial elongation factor G in rice."
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB040052; BAB13515.1; -.
 DR EMBL; AB040051; BAB13514.1; -.
 DR HSSP; P13551; 1FNM.
 DR InterPro; IPR004540; EF-G.
 DR InterPro; IPR000640; EFG_C.
 DR InterPro; IPR004161; EFTU_D2.
 DR InterPro; IPR000795; EF_GTPbind.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF00679; EFG_C; 1.
 DR Pfam; PF00009; GTP_EFTU; 1.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR PRINTS; PR00315; ELONGATINFC.
 DR TIGRPFAM; TIGR00484; EF-G; 1.
 DR TIGRPFAM; TIGR00231; small_gtp; 1.
 DR PROSITE; PS00301; EFACITOR_GTP; 1.
 KW Elongation factor; GTP-binding; Protein biosynthesis.
 SQ SEQUENCE 757 AA; 84083 MW; 93D1156FC1955182 CRC64;

Query Match 50.4%; Score 1964; DB 10; Length 757;
 Best Local Similarity 53.3%; Pred. No. 7,4e-128;
 Matches 394; Conservative 117; Mismatches 214; Indels 14; Gaps 6;

OY 3 LUGAAVVALGRGAPASLGMRKQVNMKACRWSSGVIPNEKINIGISAHIDSGKTL 62
 DB 31 LSAALAAASARGMFSASA--LRARDEKVAWRSS---MRRNIGISAHIDSGKTL 84
 OY 63 TERVLVYGRYAKMEHVKGQGVAMDSMELERQGITIOSAATFTMKVNINIDTP 122
 DB 85 TERVLVYGRYAKMEHVKGQGVAMDSMELERQGITIOSAATFTMKVNINIDTP 144
 OY 123 GHVDTIEVERALRYLDGAVLVCAVGVQCOVTMVRQMKRYNVPFLTFINKLDRMGSN 182
 DB 145 GHVDTIEVERALRYLDGAVLVCAVGVQCOVTMVRQMKRYNVPFLTFINKLDRMGAD 204
 OY 183 PARALQMRSKLNTAFTMOIPMGIEGNFKGIVDLIERATYFDDDFQGIYVGEIIPAL 242
 DB 205 PAKVNLQASKLRHNAAVQVPIGLEBEEGLVDLVELAYVFECSGGQNVVVICVPSNM 264
 OY 243 RAAATDHOELIECVANSDEOGEMLBEKIPSIDLKLAIRATLKRSFTFVPLGSAIK 302
 DB 265 QULVMEKRELIEVSEVDDDLAEFLNDEPFOANQGLAIRATYARKFETIVYMSARK 324
 OY 303 NKGVOPLLDAVLEYLPNPSEVONVAILNKDSEKKTILMNSSHHNSHPVGLAFPLEV 362

RA Yaeuda M., Tabata S.
RT "Complete genomic sequence of the filamentous nitrogen-fixing
cyanobacterium Anabaena sp. strain PCC 7120.",
RL DNA Res. 8:205-213(2001).
DR EMBL: AB003596; BAB76037.1; -
DR InterPro: IPR004540; EF-G.
DR InterPro: IPR000640; EF-G_C.
DR InterPro: IPR004161; EF-TU_D2.
DR InterPro: IPR000795; EF_GTPbind.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF006679; EF-G_C.1.
DR Pfam: PF00009; GTP_EFTU.1.
DR Pfam: PF03144; GTP_EFTU_D2.1.
DR PRINTS: PR00315; ELONGATNCT.
DR TIGRPFAMs: TIGR00484; EF-G; 1.
DR TIGRPFAMs: TIGR00231; small_GTP; 1.
DR PROSITE: PS00301; EFATOR_GTP.1.
KW Elongation factor, Complete proteome.
SQ SEQUENCE 692 AA; 76144 MW; F337ACBF03F4DD64 CRC64;

Query Match 37.8%; Score 1472; DB 16; Length 692;
Best Local Similarity 44.9%; Pred. No. 1e-93;
Matches 319; Conservative 124; Mismatches 223; Indels 44; Gaps 11;

42 PNEKINIGISAHIDSGKTTTLTTLVYTGRIAMHEVKGKDGAVGAVMSMELERQGIT 101
6 PLEKVNIGIAHIDAGKTTTTRILFYSGIHKIGEV--HEGT-AVTDMMDERERGIT 62
102 IGSAAFTMKVDNINIIDTGGVDTIEVERALRYLDGAVLVLCVGGVQCTMTVNRQ 161
63 ITAAASTSKWDQVNIIDTGGVDTIEVERSMRYLDGAVLVLCVGGVQCTMTVNRQ 122
162 MKRYNPLFTFINKLDRMGSNPARALQOMRSKLNHTAFMQIPMGLEGNGKIVDLIER 221
123 ADRYKTPRIATFNKORTGAFNRYHQRDRALRAAIAIQLPISGENDPKGIVDLIRK 182
222 AIYFDGFSQIYVYGEIPAEIAPAAATDHRQELIECVANSDEQ-----GEMPLEEKIPS 275
183 AYYMNDQGTIDTETDIPADLOQVEEYTKLVEAVETDDLSKYPQSEPLTEBEINS 242
276 ISLKAIRATLAKSFTPTPLGSAKNGVOPLLDAVLELTPNBEV-----QNYA 327
243 -----ALRGTIAGTIVPLCGSAFKNGVQMLDAVVDYLPAPPEVPEIQTTLNGDA 296
328 IINKKDSKSKTKILNNSRHNSHPVGLAFPLEVGRFGQLTVYRSYOGELKKGDITNT 387
297 IERRADNE-----PLAALAEKIMADPYGRILTVRYSGVLKKGSYVLA 341
388 RTRKVRLORLARMAHADMMEASTEVEYAGDICALFGI-DCASGDTFTDKANSGLMESIH 446
342 TKRKKERISRLVLMKADDRQ-DVEELRAGDGLAALGKDTLTGDTITDE-GAVVLESIF 399
447 VDPVYSIAMKPSNKNDLEKFSKIGRFTREDPTFKYPTTEKKEVVISGMELEHETIA 506
400 IEPVVSIVAVEPTKNDMDLSKALQSLSEBDPTFVNVDPEPTNQVIVAGMGLHEIIV 459
507 ORLEREYGCPCITGCKPVAFRETTAPV-FDFTHKQSGAGQYKVGIVLEPLPEDY 565
460 DKRLRFKFKYBANGAQVAYRETIKRVTVNVEGKFRQSGKQYGVHVINLEBGEFG-- 517
566 TKLEFDETFGSIKPROFPAVEKGLDACEKPLSGHKLSGLRFLYLDGAAHMDVSNET 625
518 TGEFYSKIVGVVPREYIGPAQMKSCESGILNGVPLIDVKATLVHGSYHVDSSSM 577
626 SFLRAEGALKALANATLCILEPTIAVEVYVAPNEGOVYIAGINRRHGVITGQDVEDY 685
578 AAFIASMSMLKEVNLVLAASPLLEPMKVEVEVEDYIGVIGDLISRRGIESQSTEQGL 637
686 FTLYADVPLNDMGSTELRSCTEGKGEYTMESRYOPLCPSQEDVINK 735
638 AKYASVPLATMNGVATDIRSKTQGRGIFTMERSHYEVPVRSVAETIINK 687

RESULT 14

OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria, Proteobacteria, epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_Taxid=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RU-1;
RA Kolesnikow T., Gekas S., Lee A.;
RT "Identification of Helicobacter pylori Antisense";
RL Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF181877; AAF04270.1; -
DR HSSP: P13551; 1DAR.
DR InterPro: IPR004540; EF-G.
DR InterPro: IPR000640; EF-G_C.
DR InterPro: IPR004161; EF-TU_D2.
DR InterPro: IPR000795; EF_GTPbind.
DR Pfam: PF006679; EF-G_C.1.
DR Pfam: PF00009; GTP_EFTU.1.
DR Pfam: PF03144; GTP_EFTU_D2.1.
DR PRINTS: PR00315; ELONGATNCT.
DR TIGRPFAMs: TIGR00484; EF-G; 1.
DR TIGRPFAMs: TIGR00231; small_GTP.1.
DR PROSITE: PS00301; EFATOR_GTP.1.
KW GTP-binding.
FT NON TER 1
SQ SEQUENCE 682 AA; 75939 MW; FC6610CC97CB584E CRC64;

Query Match 37.7%; Score 1470; DB 2; Length 682;
Best Local Similarity 45.5%; Pred. No. 1.4e-93;
Matches 315; Conservative 129; Mismatches 231; Indels 18; Gaps 11;

47 NIGISAHIDSGKTTTLTTLVYTGRIAMHEVKGKDGAVGAVMSMELERQGITQSA 106
1 NIGIAAHIDAGKTTTTRILFYSGIHKIGEV--HDG-AATMDMEQEKERGITTSNA 57
107 TTFMKVDNINIIDTGGVDTIEVERALRYLDGAVLVLCVGGVQCTMTVNRQMKRN 166
58 TTCFMDQIINLIDTGGVDTIEVERSMRYLDGAVSVFCVGGVQCTMTVNRQANKYG 117
167 VFLFTFINKLDRMGSNPARALQOMRSKLNHTAFMQIPMGLEGNGKIVDLIERAIYFD 226
118 VRIYFVNMKRIAGNPFNVENQIKLRKANVPINIGIADTIGVLDVQMAIWN 177
227 GDFSOIYRG--EIPAEIAPAAATDHRQELIECVANSDEQGLGEMPLEEKIPISIDUKLAIR 284
178 NE-TWGAQYDVEISDLEKAKYREKLVEVAQBDALEMEKYLGSELENIIEIKGK 236
285 RATLRSFTPVPLGSAKNGVOPLLDAVLELTPNBEVQNYAIIINKKDSKSKTKILMN 344
237 TGCILNMSIVPMICGSSFNKRGVOTLLDAVIDLPAPEVVDI---RGIDPTEEEVFK 292
345 SSRHNSHPFVGLAFPLEVGRF-GQLTVYRSYOGELKKGDITNTTRKKVRLORLARMA 403
293 SS--DDGEFAGIAPKIMDTPFVGQITPRYVRGKLESSYYNSTRKDKERGRLLKXHS 350
404 DMEASTEVEYAGDICALFGI-DCASGDTFTDKANSGLMESIHVPDVISIAMKPSNKN 462
351 NKRE-DIKEVYAGCICAFGLKDTLTGDTLCDEKNA-VLEBMEPEPVIHIAVEPKTKA 408
463 DLEKSKIGRTRDPTFKYPTTEKKEVVISGMELEHETIAQRLREYGCPCITGKP 522
409 DQKKGVALLKLAEDPSFRVWTOETGOTLIGGMELEHETIIVDLRERFVEAEIGQP 468

```
QY 523 KAFRETTAPVPFDTTHKKSGAGQYGVIGVLEPDLFEDYTKLEFSDETFGSNIPKQ 582
DB 469 QVAFETITSSVSKERKAKQSGRGQVHPFKLEPKKEFG--SGEYFNEISGGVTPK 526
QY 583 PVPAYEKGLDACEKGPLSGHLSGLRFLADGAAHHMVDNSRISFIRAGEGALKOALANA 642
DB 527 YLPADVKGQEMQONVNLGVIPVDPKVTLLYDSTHYDVSSEMAFKIAGSMFKESRRA 586
QY 643 TLTILEPIPAVEVVAVPAFNEFOGVYIAGINRRHGVITGQDGVEDYFTLYADVPPLNDMFGYST 702
DB 587 NVVLEPMMKVEVEVEEYMGVIGDILNRRGOINSMDELGLKYNALVPLVEMGYST 646
QY 703 ELRSCTEGKEYTMEYSRYOPLPSTQEDVINK 735
DB 647 DLRSATQGRGYTSMEDHYGEVPSNISKIIVEK 679

RESULT 15
Q9AF78 PRELIMINARY; PRT; 700 AA.
ID Q9AF78;
AC 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Elongation factor G.
GN FUS.
OS Archaeobacter sp.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Micrococcales; Micrococcaceae; Arthrobacter.
OX NCBI_Taxid=1667;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=515;
RA Berchet V., Russell N.J., Normand P., Gounot A.M.;
RT "Cloning and sequencing of the fus gene coding for elongation factor G
RL from the arctic psychrotroph Arthrobacter sp. strain 515.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ238083; CAC36321.1; -.
DR HSSP; P13551; 1DAR.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF00009; GTP_BFTU; 1.
DR PRINTS; PR00315; ELONGATINFC.
DR TIGRFAM6; TIGR00484; EF-G; 1.
DR TIGRFAM6; TIGR00231; small_GTP; 1.
DR PROSITE; PS00301; EFACITOR_GTP; 1.
KW GTP-binding.
SQ SEQUENCE 700 AA; 77478 MW; 1CC32B349330AA61 CRC64;

Query Match 36.6%; Score 1427.5; DB 2; Length 700;
Best Local Similarity 43.2%; Pred. No. 1,3e-90;
Matches 304; Conservative 134; Mismatches 240; Indels 25; Gaps 10;

QY 45 KIRNIGISAHISGKTLTERVLVYTG---RIAKGHEVKGKQGVGVAMDSMELEKRGIT 101
DB 7 KVRNIGIMAHIDAGKTTTERILFYGVNHKIGETH---DG-ASTTDMGEKKEKRGIT 60
QY 102 IGSAAFTFMKDVNINIIDTPGHVDTIEVERALRVLDGAVLVLCAGVGQCOMTMVNRQ 161
DB 61 ITSAAVTCFEMNQINIITDPGHVDFTVEVERSLRVLDGAVAVFDGKEGVEPQSEITVRQ 120
QY 162 MKRYNVPLTFLINKLDRMGSNPARALQQRSKLNHTAFMQIPMGLGNPKGIVDLIEER 221
DB 121 ADKYNVPKICFVNKMDKADAFYFVDTIISRLGVKPLVMQPLIGAENDFIVVDLLYMR 180
QY 222 AIVFDGDFSOIYRYG-----EIPALRLAATDHRQELIECVANSDEQLGEMFLEKIPS 275
DB 181 ALVWPDGSKGVDTVMGAKYEIQETPADLKEKAEYRALVETVAESSEILMEKYLEGEIT 240
```

```
QY 276 ISDLKLAIRATLKSPFVPFLGSAKKNKQVQPLLDVAVLELYLPNSSEVQNAAILNKQDS 335
DB 241 EDELKAG;RKMTINSELVPFVFGSAPKORGVQPMLDVAVDVLPLNDVPMI GHDPDBE 300
QY 336 KEKTKILNNSHHNSHPVGLAFPLEVGR-FQQLTYVASYQGLKKQDTIYTRTRKKYR 394
DB 301 KELTR-----KPSSEPPSALAFKIATHPFQQLTFIRVYSGHVGAGQVAVNSTGKKER 355
QY 395 LQRLARHMDMWEASTEERYVAGDICALFGI--DCAGDPFTDKNAGSLMESIHVDDPYIS 453
DB 356 IGLKQTMHAN-KEMVVDGATGHYIAALGLKDTTGGTLCBPANO-IYLSHSPFEPVVIS 413
QY 454 IAKPSSNKNDLEKFSKGIGRFTREDPTKVVFDTENKSTVYISGMGLHLIETVQRLEREY 513
DB 414 VAIEPNTGDDQEKSTALQKLSAEDPTFOVSLNEDTGTITAGMGLHLIDLIVDMRRERF 473
QY 514 GCPCTIGKPKVAFRETTIAPVP-FPETHKKSGAGQYGVIGVLEPDLFEDYTKLEFSD 572
DB 474 KVEANVGRPOVAYRETIKRAVERHDYTHKKQYSGSGQPAKIQIAIEPLDTSSEGLYERPN 533
QY 573 ETPGSNIPKQFVPAYEKGPLDACEKGPLSGHLSGLRFLADGAAHHMVDNSRISFIRAGE 632
DB 534 KVTGRIPEEYIPSVDAQIQDLANDVLAGIPVVGIKATLIDGAYHVDSSMARKIAGR 593
QY 633 GALKOALANATLCILEPIPAVEVVAVPAFNEFOGVYIAGINRRHGVITGQDGVEDYFTLYADY 692
DB 594 MAFKRAARKANPILLEPLMDVEVTRPEYMGVIGDILNRRGOINSMDELGLKYNALVPLVEMGYST 653
QY 693 PLNDMFGISTELRSGTEGKEYTMEYSRYOPLPSTQEDVINK 735
DB 654 PLSGMFGYIGDLRSKTQGRAVVSMTHSYAEVFKAFADIIQK 696
```

Search completed: July 14, 2003, 18:21:01
JDB time : 39.8672 secs

THIS PAGE BLANK (USPTO)

FT NP_BIND 175 178 GTP (BY SIMILARITY).
 SQ SEQUENCE 752 AA; 83770 MW; DFB6108D38A72E4 CRC64;
 Query Match 83.2%; Score 3243; DB 1; Length 752;
 Best Local Similarity 84.1%; Pred. No. 1.7e-206;
 Matches 636; Conservative 37; Mismatches 75; Indels 8; Gaps 5;

QY 1 MRLGAAVAALGRG--RAPASLGWQRKQVNMKACRWSSGGVLPNEKIRNIGISAHIDS 57
 DB 1 MRL--RTFAGLGRPLPRVPALILGWQGANWKTIRMWSSGSIPIKEKIRNIGISAHIDS 58
 QY 58 GKTTLTERVLYYGRGAKMHEVKGKGVAVNDMSLEEROGITTSATFTMKGVNIN 117
 DB 59 GKTTLTERVLYYGRGAKMHEVKGKGVAVNDMSLEEROGITTSATFTMKGVNIN 118
 QY 118 IIDTPGHVPTIEVERALRLDGAIVLVCAVGVCQMTVNNQMRVYVPLTFINKD 177
 DB 119 IIDTPGHVPTIEVERALRLDGAIVLVCAVGVCQMTVNNQMRVYVPLTFINKD 178
 QY 178 RMGSNPARALQOMRSKLTNHTAFMOIPMGIEGNFKGIIVDLIERATYFDGFSQIVRYGE 237
 DB 179 RMGSNPARALQOMRSKLTNHTAFMOIPMGIEGNFKGIIVDLIERATYFDGFSQIVRYGE 238
 QY 238 IPBBLAATDNRQSLIECVANSDEOLGEMFLEEKIPISIDKLAIIRATLKRSFTPVFL 297
 DB 239 IPADLAAADNRPELIECVANSDEOLGEMFLEEKIPISVSDKLAIIRATLKRSFTPVFL 298
 QY 298 GSALKNKGVOPLDAVLEVLNPSEVONYALINKDSKEKTKILNNSRNHNPVGLA 357
 DB 299 GSALKNKGVOPLDAVLEVLNPSEVONYALINKDSKEKTKILNNSRNHNPVGLA 357
 QY 358 FPLEVGRFGQLYVRSYOGELKKGGDTYNTTRKKVRLORLARMAHDMMEASTEVEYACD 417
 DB 358 FPLEVGRFGQLYVRSYOGELKKGGDTYNTTRKKVRLORLARMAHDMMEASTEVEYACD 416
 QY 418 ICLFETIDASGDTFTDKANSGLSMESINHPDVIISANKPSKNDLEKFSKIGFTPE 477
 DB 417 ICLFETIDASGDTFTDKANSGLSMESINHPDVIISANKPSKNDLEKFSKIGFTPE 476
 QY 478 DPTFKYVFTDENETVYISGMGELHLEIYAQRLEREGPCITGPKVAVRETTTAVPPD 537
 DB 477 DPTFKYVFTDENETVYISGMGELHLEIYAQRLEREGPCITGPKVAVRETTTAVPPD 536
 QY 538 FTHKKSOGGA-GQYKVGIVLEPLDEDTYKLEFSDTEGSGNIPKQFPAVEKGFPLDACE 596
 DB 537 ITHIKSSRVVPASMGKIVGLPELPAEDLPKLEFSDTEGSGNIPKQFPAVEKGFPLDACE 596
 QY 597 KGPLSGHKLSGLRFPVODGAHHNVDSNEISFTIRAGGALKQALASATLCIIEPIMSVEVI 656
 DB 597 KGPLSGHKLSGLRFPVODGAHHNVDSNEISFTIRAGGALKQALASATLCIIEPIMSVEVI 656
 QY 657 APPEFGOVLAGINRRHGVITGQGVEDFTLVADVPPLNDMPGYSTELSCTEGKEGYTM 716
 DB 657 APPEFGOVLAGINRRHGVITGQGVEDFTLVADVPPLNDMPGYSTELSCTEGKEGYTM 716
 QY 717 EYSRYOPLPSTQSDVINKLEATGQLPVKGRKAKN 752
 DB 717 EYSRYOPLPSTQSDVINKLEATGQLPVKGRKAKN 752
 RESULT 2
 EFGI_YEAST STANDARD; PRT; 761 AA.
 AC P25039; Q99360; 01-MAY-1992 (Rel. 22, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DE 15-JUL-1999 (Rel. 38, Last annotation update)
 GN MEPI OR YLR069C OR L2195. mitochondrial precursor (MEP-G-1).
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycaceae;
 NCBI_TaxId=4932;

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92037620; PubMed=1935960;
 RT Vamvutae A., Ackeman S.H., Tzagoloff A.;
 RT "Mitochondrial translational-initiation and elongation factors in
 RT Saccharomyces cerevisiae";
 RL Eur. J. Biochem. 201;643-652(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Andre B., Ureastarazu L.A.;
 RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
 CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
 CC RIBOSOME.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC EF-G/EF-2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; X58378; CAA41267.1; -;
 CC EMBL; X94607; CAA64315.1; -;
 CC EMBL; Z73241; CAA97626.1; -;
 CC PIR; S17025; S17025.
 CC PIR; S20179; S20179.
 CC HSP; P13551; IFFM.
 CC SGD; S0004059; MEPI.
 CC InterPro: IPR004540; EF-G.
 CC InterPro: IPR000640; EFG C.
 CC InterPro: IPR004161; EFTU D2.
 CC InterPro: IPR000795; EF_GTPbind.
 CC InterPro: IPR005225; Sma11 GTP.
 CC Pfam; PF00009; GTP_EFTU; 1.
 CC Pfam; PF00679; EFG_C; 1.
 CC Pfam; PF03144; GTP_EFTU D2; 1.
 CC PRINTS; PR00315; ELONGATNFACT.
 CC TIGRfam; TIGR00231; sma11_GTP; 1.
 CC TIGRfam; TIGR00484; EF-G; 1.
 CC PROSITE; PS00301; EFACITOR GTP; 1.
 CC Elongation factor; Protein biosynthesis; Mitochondrion;
 CC Transist peptide; GTP-binding.
 CC TRANSIT 1
 CC CHAIN ?
 CC NP_BIND 77 761
 CC NP_BIND 148 84
 CC NP_BIND 152 152
 CC NP_BIND 202 205
 CC CONFLICT 66 66
 CC CONFLICT 233 233
 CC CONFLICT 478 478
 CC CONFLICT 629 629
 CC SEQUENCE 761 AA; 84573 MW; 3E2C534509B09103 CRC64;
 Query Match 51.0%; Score 1987; DB 1; Length 761;
 Best Local Similarity 56.2%; Pred. No. 1.4e-123;
 Matches 392; Conservative 117; Mismatches 172; Indels 16; Gaps 8;

QY 45 KIRNIGISAHIDSGKTTTLTERVLYYGRGAKMHEVKGKGVAVNDMSLEEROGITTS 104
 DB 69 KIRNIGISAHIDSGKTTTLTERVLYYGRGAKMHEVKGKGVAVNDMSLEEROGITTS 128
 QY 105 AATFTMW---KDVNINIIDTPGHVPTIEVERALRLDGAIVLVCAVGVCQMTVNR 160
 DB 129 AATFTMWKEGKNYHFNIDTPGHVPTIEVERALRLDGAIVLVCAVGVCQMTVNR 188
 QY 161 QMRNVNVPPLTFINKDMSGNPARALQOMRSKLTNHTAFMOIPMGIEGNFKGIYDLIEE 220
 DB 189 QMRNVNVPPLTFINKDMSGNPARALQOMRSKLTNHTAFMOIPMGIEGNFKGIYDLINR 248

Db 595 VVLEPIRVSVEGPOEFGSVFGLINRGVAVSSADDEGFSVVDALVPLESEKFGSTVL 654
QY 705 RSTCEKGETMYEYSRQPCLPSTQEDVINKYLE 738
Db 655 RSTQGAKEYSMEFARKAPQGVTDLSLKEYDE 688

RESULT 4
EFG1_BORBU STANDARD; PRT; 693 AA.
AC 030913; 051490;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor G 1 (EF-G 1).
GN FUSA OR FUS1 OR BB0540.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxId=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DK1;
RA Sonderbye L., Hindersson P., Pedersen S.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RA MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Karlyaveva A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Utecherback T., Wachtel L., McDonald L., Arlisch P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi."
RL Nature 390:580-586(1997).

CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
RIBOSOME.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; AF021260; AAB71893.1; -
CC EMBL; AE001155; AAC68897.1; -
CC HSSP; P13551; 1ELO.
DR TIGR; BB0540;
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR TIGRfam; TIGR00231; small_GTP; 1.
DR TIGRfam; TIGR00484; EF-G_1.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KM Elongation factor; Protein biosynthesis; GTP-binding;
FT NP_BIND 13 20 GTP (BY SIMILARITY).

FT NP_BIND 80 84 GTP (BY SIMILARITY).
FT NP_BIND 134 137 GTP (BY SIMILARITY).
FT CONFLICT 201 201 D -> E (IN REF. 1).
FT CONFLICT 208 208 S -> N (IN REF. 1).
FT CONFLICT 212 212 I -> M (IN REF. 1).
FT CONFLICT 216 216 T -> A (IN REF. 1).
FT CONFLICT 236 236 T -> I (IN REF. 1).
FT CONFLICT 243 243 T -> I (IN REF. 1).
FT CONFLICT 304 304 E -> S (IN REF. 1).
FT CONFLICT 331 331 I -> T (IN REF. 1).
FT CONFLICT 331 331 I -> G (IN REF. 1).
FT CONFLICT 484 484 R -> A (IN REF. 1).
FT CONFLICT 504 504 D -> A (IN REF. 1).
FT CONFLICT 513 513 S -> N (IN REF. 1).
FT CONFLICT 548 548 R -> K (IN REF. 1).
FT CONFLICT 671 671 R -> K (IN REF. 1).
FT CONFLICT 678 678 T -> A (IN REF. 1).
FT CONFLICT 689 689 D -> R (IN REF. 1).
SQ SEQUENCE 693 AA; 7572 MW; F8A44C38D58F9B1E CRC64;

Query Match 45.2%; Score 1762; DB 1; Length 693;
Best Local Similarity 51.8%; Pred. No. 8.9e-109;
Matches 361; Conservative 126; Mismatches 188; Indels 22; Gaps 10;

QY 45 KIRNIGISAHIDSGKTTLTERTVLYTGRJAKNHEVKGXGQGVAVNDSELERQGITIOS 104
Db 5 KLRNIGISAHIDSGKTTLTERTLIFYCNKIHAIHEVKGXGQGVATWDSMELEREGITIAS 64

QY 105 AATFTMKDVNINIIDTDEGHVDFTEVERALRVLDGAVLYLCAVGVQCCMTNVRQMR 164
Db 65 AATHWEKDFPNIIDTDEGHVDFTEVERSLRVLDGALIVDSVAGVQSISTVRQLR 124

QY 165 YVVPFLTFPINKDRMGSPARALQOMRSKLNNTAFMOIPMGLNFGKIVDLIEBAY 224
Db 125 YVPLPLAFNKKDKRGANPVNKKDRLKSLNSVLMQPILEDKHGIVDLVLMKAY 184

QY 225 PDG-DPSQIVRYGEIPABIRAAATDRQELIECVANSBQLEMFLEEKIPISDLKAI 283
Db 185 FPGKDGTEIIE-KEIPDLIEAKSKREIMDTLADFDELMEHMEKEVPEITIIYNT 243

QY 284 RAAITLKRSFTVPLGSAALKNGVQPLDAVLEYLNPSPVQVYALINKDSEKETKILM 343
Db 244 RTGTIALKLCVPMFSAYKNKGVQLLDVAVRFLSPDHDKNTALDNNNEIKDLK-- 301

QY 344 NSGRNHSHPFVGLAPPLEVGRFQGLTYVRSYSGELKGGTYNTRRKKVRLQRLARMA 403
Db 302 ---DNEPFTVALAKLEDGQYGLTYRITVIGILAKGOELNSRKSXKGRILIRMA 357

QY 404 DMEASTSEEVY---AGDICALFGIDCASGDTFTDKANSGLSNEISHVDPVISIAMPSN 460
Db 358 N---NTEDIEFGSGDIVALFGICASGDTFCD-PSINYSWTSMFIPDPVISLSVKPYD 412

QY 461 KNDLEKFSIGIRFTRDEPTFPVYPTENKEVIVSGMELHLEIYAQRLERYGRCPTG 520
Db 413 KKSADNMMAKALGRFKEDPTFTYVDISNETITIGMELHLEVIERKREFKAEVETG 472

QY 521 KPVAFRETTAPVFPDFTHKKQSGAGQYGVIVLEPLDPEBDTKLEFSDETGSNP 580
Db 473 MPQVAYRETTAKAPNTTHKKQSGAGQFGVAFMBELDGEET-YEFVVLIGGVYIP 531

QY 581 KQFVAVKEXGLDACERKPLSGHKLISGLRPVLDGAHHNVDSNEISFRAGEALKQALA 640
Db 532 TEYIPSCDKGFQKAMERGLTIGFPIVDIKITINDQYHIVDSSDIAFOLAIGAFREAYE 591

QY 641 NATLCILPEYMAVEVVAENRPGQYIAGINRHGVITG--QDGVADFTLVADVPLNMF 698
Db 592 KAKPTLIEPIMKVTLEGEFEGNMFGILNQRGIITSLBEG--SFSKVEAEVPLESEMF 649

QY 699 GYSTELEKSTCEKGETMYEYSRQPCLPSTQEDVINK 735
Db 650 GFSTVLRSTQGAKEYSMEFARKAPQGVTDLSLKEYDE 685

RESULT 5

```

EFG_SYNP6
ID EFG_SYNP6 STANDARD; PRT; 694 AA.
AC P18667;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor G (EF-G).
GN FUS1 OR FUS.
OS Synechococcus sp. (strain PCC 6301) (Anacystis nidulans).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
NCBI_TaxID=1139;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=89281486; PubMed=2499762;
RA Meng B.-Y., Shinozaki K., Sugita M.;
RT "Genes for the ribosomal proteins S12 and S7 and elongation factors
RT EF-G and EF-Tu of the cyanobacterium, Anacystis nidulans: structural
RT homology between 16S rRNA and S7 mRNA."
RL Mol. Genet. 216:25-30(1989).
CC -! FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -! SUBCELLULAR LOCATION: Cytoplasmic.
CC -! SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X17442; CAA55495.1; -
DR PIR: S04429; S04429.
DR HSP: P13551; 2EEG.
DR InterPro: IPR0004540; EF-G.
DR InterPro: IPR000640; EF-G.
DR InterPro: IPR004161; EFTU D2.
DR InterPro: IPR000785; EF-GTPbind.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00009; GTP_EFTU; 1.
DR Pfam: PF00679; EFG_C; 1.
DR Pfam: PF03144; GTP_EFTU D2; 1.
DR TIGRPFAM: TIGR00231; small_GTP; 1.
DR TIGRPFAM: TIGR00484; EF-G; 1.
DR PROSITE: PS00301; EFATOR_GTP; 1.
DR Elongation factor; Protein biosynthesis; GTP-binding.
KW NP BIND
FT NP BIND 17 24 GTP (BY SIMILARITY).
FT NP BIND 81 85 GTP (BY SIMILARITY).
FT NP BIND 135 138 GTP (BY SIMILARITY).
SQ SEQUENCE 694 AA; 75909 MW; D1ADFCTD515C95FF CRC64;
Query Match 38.3%; Score 1493; DB 1; Length 694;
Best Local Similarity 44.5%; Pred. No. 5, 1e-91;
Matches 315; Conservative 128; Mismatches 229; Indels 36; Gaps 9;
QY 41 INNEKIRNIGISAHIDSGKTLTERLYTTGRIAKHEVYKGDGVAVMDSELERQCI 100
DB 5 VLEKVRNIGIAHIDAGKTTTERTILFYSGVHKIGEV--HDG--NAVVMWMEOEEREI 61
QY 101 TIOSAATFTMKDQVNIINIDTPEGHVFTIEVERALRVLGAVLVICAVGVCCQMTTVAR 160
DB 62 TTTAAHISTSMKDYRNIIIDTPEGHVFTIEVERSMKVLGVAVAFCSVGQVQSTTVAR 121
QY 161 QMKRYVNPFLTRINKLDRKSNPARALQQRSKLNTNTAFMOIPLMGLEGNFKGIVDLIE 220
DB 122 QADRVSVPRIIVFNKKDRGADDFKYYGDIRVRANAVPIQIPGIAESDPFGIVDLVEM 181
QY 221 RALYFPGDPSQIVRYVEIPAELEAAATDRQELIECVANSDEQGLGEMPLEEKIPISDIK 280
DB 182 KAHITNDIGTDILVTIDPAELQETAAEWSKRVAVAEETDEALDKYFEDDLSIEDIK 241

```

```

QY 281 LAIRRATLKRS---FTPVFLGSLNKKGVOPLLDAVLEYPNPSEVQ-----NYAI 328
DB 242 AGLRKGVLLIQGDRVLVPLVPCSAFNKGVQQLLDNAVLELPSRPOIIPPIQGLPGGVALL 301
QY 329 LNKKDSKREKTKILMNSSRHNSHPFVGLAFPLEVGRFQGLTVRSYQGLKKGGDTYNTN 388
DB 302 RPSDEA-----PSSALAFKIMADPYGRITFVRVYSGILQKGSYYVNT 345
QY 369 TRKTRRLQRLARMHADMMEASTEEVYAGDICALPOI--CCASGDTPTDANKSLMESIHV 447
DB 346 KKKKRVSRLLIILKAD-DRIEYDELRAQDLGAVLGLKPTGDTLCDQNP--IILSLPT 403
QY 448 PDPVISIMKPSNKNKDLLEKFSKIGRFTREDPTFKVYEDTEKTEYISGMGLHLEIYAO 507
DB 404 PEPVSAVNEPRTKNDMEKLSALQALSEBDPTFVSVDSERNQVIVMGSLHLEIYLD 463
QY 508 RLEREYGCPCITGKRKVAPRETTIAPVPDPPTHKKQSGAGQYGVIVGLPELDPEDYTK 567
DB 464 RMLREYKVEANIGADQVAVRETVRAVYAEKGFVRSQSGKQYGHVIELEBPAPG--TG 521
QY 568 LEFSDETGSNIPKQFVAVKGFPLDACEKGPLSGHKLGLAFVLDGAAHNVDSNEISF 627
DB 522 FEFVSKIVGTVPKSYVGPAPKQKTECSGVLAGYPLIDIKATLVDSYHDVDSSEMAF 581
QY 628 IRAGEGALKQALANATLILPEIMAVEVAPNEPFGQYIAGINRRHGYITGQGVEDYPT 687
DB 582 KTAGSMATKEAVRKADPVLLEVMVVEVPEDFISVMGNLISRQILEGATTINGRT 641
QY 668 LVADVPVLMDFGYSSTELNSCTEGKGYTMEYSRYPCLPSTGEDVYINK 735
DB 642 VSAKVPFLAMEFGYATDLSMTGSGRIPTMEFSQYEVPRNAVETIIAK 689
RESULT 6
EFG_HELP
ID EFG_HELP STANDARD; PRT; 691 AA.
AC Q9ZK24;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor G (EF-G).
GN FUS1 OR JHP1118.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
NCBI_TaxID=85963;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moll D.T.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori."
RL Nature 397:176-180(1999).
CC -! FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -! SUBCELLULAR LOCATION: Cytoplasmic.
CC -! SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

Query Match	38.2%;	Score 1489;	DB 1;	Length 691;
Best Local Similarity	45.7%;	Pred. No. 9.2e-91;		
Matches 319;	Conservative 128;	Mismatches 233;	Indels 18;	Gaps 11;

Qy	102	OSATFPMKDDVANI-IDTGHVFTLEVEALRVLBGALVYLCAVGVGOQOMTVARQ	161
Dd	62	ITSATITCFMKDHOJNLIDTGHVFTLEVEALRVLBGALVYLCAVGVGOQOMTVARQ	121
Qy	162	MKRYNVPFLRTJINKLDWGSNPARALOQMSKLNNHTA FMOJ PMGLBGNFGJVDLIEER	221
Dd	122	ANKRGVPRIVEVNNKMDRGANFYVENVQIKQRLKANPVPINIPICABDTFIGVIDLVQMK	181
Qy	222	AIYDGDSSQVRYG--EIPALRPAALDTHOELIECYANSDQOLGEVFLBEKIPSIDL	279
Dd	182	AIWNNE--TMGAKYDVEEIPSDLLEKAKOYERKEVEAAEDDEALMEKYLIGBEBLIEEI	240
Qy	280	KLAIIRATILKESFTFVFLGSAJLKNKGVOPLDVALEYVNPSEVONTAIIINKDDSEKT	339
Dd	241	KKGIKTGGLNMSFVVMLCGSSFKKKGVOTLLDAVIDIYAPLPEVDI----KGIDPKTEE	296
Qy	340	KILNNSHNHSPVGLAFPLEVGRF--GQTYVRSYOGELKKDITTYNTRKKVRLQRL	398
Dd	297	EVFVYSS--DDGEFAGLAFKIMTDPFVQGLFVRYRRGLKESGSVYVNSTDKKERVRL	354
Qy	399	ARMHADMMWASTBEVYAGDICALGEI--DCASGDPTDKANGSLMESIHVDDPVYSIMK	457
Dd	355	LKMSNKRK--DIKEYTAGELCAFVGLKOTLLTGDTLDCDKMA--VLLERMEFPFVTHIVE	412
Qy	458	PSNRKDLKFSKJIGRFTRDEPTFKYVDTENKETVISGMBLHLEIYAORLEREYGCPC	517
Dd	413	EKTRADQCKMGVAGLKLAEBDPSFVVMVQBEICQTLIGSGWELHLEIIVDLRKRFXYEA	472
Qy	518	ITGRPKYAFRETTIAPVPFDTTHKKQSGAGQYKVIQVLEPLDEBDYTKLFSDETGS	577
Dd	473	EIGQPOVAFRETIIRSVSKEHKYAAQSGRGQYGVHFTLBPKEBG--SGYEFVNEIISG	530
Qy	578	NI PKQFVAVEKGFIDACKEKPLSGHKSGLRFLVODAHMVBNSIEFIRAGGALQO	637
Dd	531	VIPKEYIYAVDKGIOEMQNGVLAGYPAVDEKVTLLYDOSYHDVDSSEMAFKAIGSMATKE	590
Qy	638	ALANATLICILEPIAWEVVAENEFQGOVIAGINRRHGVITQDQVEDFTLYADVPLMD	697
Dd	591	ASRAANPVLLEPPMKVEVEVEEYMGVIGIDINRRRGINSMDRLGLKIVNAPVPLEM	650
Qy	698	FGYSTELRSCTEBGGEYMETYSRQOCLPSPQEDVYINK	735
Dd	651	FGYSTDLNSATQGRGTYSMEBDHGVSEVNSIAKELVEK	688

```

RESULT 7
EFG_HELPY
ID EFG_HELPY STANDARD; PRI; 691 AA.
AC P56002;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor G (EF-G).
GN FUSA OR HP1195.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
CC Helicobacter.
XX NCBI_TaxId=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STPAIN-26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klein H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Richardson J., Zhou L., Kirkness E.F., Peterson S.,
RA Lettun B., Quackenbush J., Zhou L., Kirkness E.F., Glodex A.,
RA McEnney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Cockayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Goddard J.M., Fujii C., Bowman C., Wathey L., Wallin E.,
RA Hayes W.S., Bordovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
CC RIBOSOME.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its use by
CC non-profit institutions as long as its content is in no way used for
CC commercial or other purposes. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
-----
CC --
DR EMBL; AE000625; AACD08239.1; -.
DR HSSP; P13551; IELO.
DR TIGR; HP1195; -.
DR InterPro; IPR004540; EF-G.
DR InterPro; IPR000640; ERG_C.
DR InterPro; IPR004161; EFTU_D2.
DR InterPro; IPR000795; EF_GTPbind.
DR InterPro; IPR005225; Small_GTP.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF00679; ERG_C; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATINFCT.
DR TIGRFAMs; TIGR00231; small_GTP; 1.
DR TIGRFAMs; TIGR00484; EF-G; 1.
DR PROSITE; PS00301; EFACFOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.
FT INT_MET 0 BY SIMILARITY
FT NP_BIND 16 GTP (BY SIMILARITY)
FT NP_BIND 80 GTP (BY SIMILARITY)
FT NP_BIND 134 GTP (BY SIMILARITY)
SQ SEQUENCE 691 AA; 76890 MW; C9SC505FC2D22667 CRC64;
Query Match 38.1%; Score 1487; DB 1; Length 691;
Beat Local Similarity 45.7%; Pred. No. 1,3e-90;
Matches 319; Conservative 129; Mismatches 232; Indels 18; Gaps 11;

```

QY 42 PNEKIRNIGISAHIDSGKTTLTERLVYTGRIAKNEHVKDGVGAVNDMELEORGIT 101
 DB 5 PLNRIRINIGIAAHIDAGKTTTERRILFYGVSHKIGEV--HDG-AATMDMMEQEKERGIT 61
 QY 102 IOSAATFTMVKDVINIIDTPGHVDFTIEVERARLVLDGAVLVLCVGVGVOGQMTVNRQ 161
 DB 62 ITSAAITCWMKHQJNLIDTPGHVDFTIEVERSMRVLDGAVVFCVGVGVQOSERVNRQ 121
 QY 162 MKRYVVPFLFTINKLDRMGSNPARALQOMRSKLNHTAFMOJPMGLEGNFKGIIVDLIEER 221
 DB 122 ANKGVPRIVFNKMDRIGANPNVNZQIKALKANPVEKXNPIGMEDEFFIGVIDLVQMK 181
 QY 222 AIYFGDSQIVRYG--EIPAEIRAAATDHRQELIECVANSBQUGEMLEEKIPISIDL 279
 DB 182 AIYVNNNE-TWGAKEYDVEIIPSDLLEKAKEREKLVAVAEQDBALMEKYLGESELEIERI 240
 QY 280 KLAIRATIKRSFTPLFGSALKNGVPLDVALEYLPNPSVQNYALINKKDSKERT 339
 DB 241 KKGIRAGCLNMLVPLCGSSFFKNKGVQTLDAVIDLYAPTEVNDI---KGIDPKTER 296
 QY 340 KILNNSRHNHPVGLAPFLVGRF-GQLTYVRSYQGEIKKGDITYNTRTKRYLQRL 398
 DB 297 EYFVSS--DDBEPFAGLAKIMTDPVQGLTFVRYRGKLESGSYVNSTKXKERVGR 354
 QY 399 ARWADMEASTEEVYAGDICALFGI-DCASGDTFDKANSGLSMESIHVPDVISIAWK 457
 DB 355 LKMSHNSKRE-DIKEYVYAGEICAFVGLKDTLTGTLCDERNA-VLERMEFPPEVHIAYE 412
 QY 458 PENKNDLEKFSKIGRFTEDPFPKRYPTTENKETYISNGELHLEIYAQRLEREKRGCC 517
 DB 413 PTKVADQEKMGVALGLAEEDPSFRVWTOEETGQTLIGMGEHLHEIYDRKREPKYA 472
 QY 518 ITGKRPVARETITTAVPFDPFTHKQSGAGQYKIVGLLEPLDPEDYTKLEFSDETFSS 577
 DB 473 EIGQOVARETIRRSVSKENKAKNSGGRGOYGHVPILEKREPS--SGYFVNEISG 530
 QY 578 NIPKQFVAVEKGFLDACEKGPLSGHKLSGRLVLDGAAHNVDSNEISFIRAGECALQ 637
 DB 531 VIPKEVTPAVDKDIGRAMONGVLAVGVDPFKVTLIDGSHYDVSSEMAFKLAGMAFKE 590
 QY 638 ALANATLCTLEPIMAVEVVAPEFQGVYAGINRRRGVITGQDGVEDYTLVADVPNDM 697
 DB 591 ASRAANPVLLPEPMKKEVEVPEEYMGDVIGDINRRGQINSMDRLGLKTIKIVAFPLVEM 650
 QY 698 FGYSTLESCTEGKEGYTMEYRYQCPLESTOEDVINK 735
 DB 651 FGYSTDLRSATOGKTYSMEPDHYGEVPSNIAKEIYEV 688

RESULT 8
 EFG_YERPE STANDARD; PRT; 702 AA.
 AC 0827B3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Elongation factor G (EF-G).
 GN FUS1 OR YPO0202.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxId=632;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=CO-92 / Biovar Orientalis;
 MEDLINE=21470413; PubMed=11586360;
 Parikh J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebatina M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cardeno-Tarraga A.M.,
 RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Felwell T., Hamlin N., Holtroyd S., Uggels K., Kariyeh A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.;

RT "Genome sequence of Yersinia pestis, the causative agent of plague.";
 RL Nature 413:523-527(2001).
 CC -1- FUNCTION: This protein promotes the GTP-dependent translocation of
 CC the nascent protein chain from the A-site to the P-site of the
 CC ribosome.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC EF-G/EF-2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL, AJ414141; CAC99063.1; -
 CC InterPro; IPR004540; EF-G.
 CC InterPro; IPR000640; EF-G.
 CC InterPro; IPR004161; EFTU D2.
 CC InterPro; IPR000795; EF_GTPbind.
 CC InterPro; IPR005225; Small_GTP.
 CC Pfam; PF00679; EF-G; 1.
 CC Pfam; PF00009; GTP_EFTU; 1.
 CC Pfam; PF03144; GTP_EFTU D2; 1.
 CC PRINTS; PR00315; ELONGATNFCT.
 CC TIGRPFams; TIGR00484; EF-G; 1.
 CC TIGRPFams; TIGR00231; small_GTP; 1.
 CC DR PROSITE; PS00301; EFACITOR_GTP; 1.
 CC KW Elongation factor; Protein biosynthesis; GTP-binding;
 CC Complete proteome.
 CC FT NP_BIND 17 24 GTP (BY SIMILARITY).
 CC FT NP_BIND 88 92 GTP (BY SIMILARITY).
 CC FT NP_BIND 142 145 GTP (BY SIMILARITY).
 CC SQ SEQUENCE 702 AA; 77537 MW; 8977FC9516165D2A CRC64;

Query Match 37.5%; Score 1462; DB 1; Length 702;
 Best Local Similarity 44.4%; Pred. No. 5,7e-89;
 Matches 317; Conservative 132; Mismatches 235; Indels 30; Gaps 15;

QY 42 PNEKIRNIGISAHIDSGKTTLTERLVYTGRIAKNEHVKDGVGAVNDMELEORGIT 101
 DB 6 PLERYRINIGISAHIDAGKTTTERRILFYGVNHKIGEV--HDG-AATMDMMEQEKERGIT 62
 QY 102 IOSAATFTMVKDVINIIDTPGHVDFTIEVERARLVLDGAVLVLCVGVGVOGQMTVNRQ 154
 DB 63 ITSAAITCWMKHQJNLIDTPGHVDFTIEVERSMRVLDGAVVYFCVGVGVQOSERVNRQ 122
 QY 155 TMTVNRQKRYNVPFLFTINKLDRMGSNPARALQOMRSKLNHTAFMOJPMGLEGNFKGI 214
 DB 123 SETVNRQANKRYKVPRIAVNKNDRGANFLVVGQKSRKANPVPDLAIGAEKFPNGI 182
 QY 215 VDLIEERAI-YFDGDSQIVRYGGEIPAEIRAAATDHRQELIECVANSBQUGEMLEEKI 273
 DB 183 IDLVGMKAIINNNDAGQTFEYBEIPAMAMELAEMHQNVLVSAAABADBLMDKYLGESE 242
 QY 274 PSISDLKLAIRATIKRSFTPLFGSALKNGVPLDVALEYLPNPSVQNYALINKKD 333
 DB 243 LTBEEKIKALKORVLKSEIILVTCGSAFKNGVQAMDAVLEYLPAPPDVE--SINGILD 300
 QY 334 DSKERTKILNNSSRH--NSHPFVGLAPFLVGRF-GQLTYVRSYQGEIKKGDITYNTRTR 390
 DB 301 DGDQTPAV-----RHSDKEPFSALEKIALNDPFGNLTFRVYSGIVNSGDTVANSYKS 355
 QY 391 KKVRLQRLARNHADMEASTEEVYAGDICALFGI-DCASGDTFDKANSGLSMESIHVPD 449
 DB 356 QREIRGRIVQNHANRRE-EIKEVYAGDIAAIGLKDVTGTGLCD-PNNPIILERMEPPE 413
 QY 450 PYISIAKMPKNDLEKFSKIGRFTEDPFPKRYPTTENKETYISNGELHLEIYAQRL 509
 DB 414 PVISAVVBPKTQADQEKMGVALGLAEKEDPSFRVWTOEETGQTLIGMGEHLHEIYDRM 473

QY 510 EREYGGPCITGKPKVAFRETTITAPV-PFDTHKKQSGAGAGYGVIGVLEPLDPEDYTKL 568
DB 474 RREFFVEANVGKQVAVYRETIKVDKVEKHAQSGRGQYGHVYIDMSPLPPG-VG 532
QY 569 EFSEDFEFGSINIPKQVPAVEKGLDACEGKPLSGHLSGLRFLVLDGAAHMDNSNISFI 628
DB 533 EFNVEIVGGISIPKFIKPAVDKGIQEQAKSGPLAGYVVDKVLHAGSYHADVSSBLARK 592
QY 629 RAGEGALKQALNATATCILEPIMAVEVAPNEFOGVIAGINRRHGVITGQDGVEDYFTL 688
DB 593 LAGSIAFKBGFKAQKAVLEPIKMEVEETPEDYMGVMDLNRRIIEGMEETATGKTV 652
QY 689 YADVPLNDMGYSTELRSCTEGKGEYTMESRYOPLPSTQEDVYINKYLEATQ 742
DB 653 RVKVPLESMGCVATDLRSQGRASYSMEFLVABA-PS--NVAKAVIEARCK 702

RESULT 9
EFG_CAMJ2
ID EFG_CAMJ2 STANDARD; PRT; 691 AA.
AC 09PIL6;
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Elongation factor G (EF-G).
GN FUSA OR CJO493.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxId=197;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Barkhill J., Wren B.W., Mungell K., Kestley J.M., Churcher C.,
Baehum D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
Jagels K., Kariyeh A.V., Moul S., Pallen M.J., Penn C.W.,
Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
Whitehead S., Barrett B.G.;
RA "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
CC -1- FUNCTION: This protein promotes the GTP-dependent translocation of
CC the nascent protein chain from the A-site to the P-site of the
CC ribosome.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-G/EF-2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL139075; CAB75131.1; -.
DR HSSP: P13551; 1ELO.
DR InterPro: IPR004540; EF-G.
DR InterPro: IPR000640; EFG_C.
DR InterPro: IPR004161; EFTU_D2.
DR InterPro: IPR000795; EF_GTPbind.
DR InterPro: IPR005225; Small_GTP.
DR Pfam: PF00679; EFG_C.1.
DR Pfam: PF00009; GTP_EFTU.1.
DR Pfam: PF03144; GTP_EFTU_D2.1.
DR PRINTS: PR00315; ELONGATINFC.
DR TIGRFAWS: TIGR00484; EF-G.1.
DR TIGRFAWS: TIGR00231; small_GTP.1.
DR PROSITE: PS00301; EFATOR_GTP.1.
KW Elongation factor; Protein biosynthesis; GTP-binding;
KW Complete proteome.

FT NP BIND 17 24 GTP (BY SIMILARITY).
FT NP BIND 81 85 GTP (BY SIMILARITY).
FT NP BIND 135 138 GTP (BY SIMILARITY).
SQ SEQUENCE 691 AA; 76719 MW; 16D3BF62B749CC60 CRC64;
Query Match 37.5%; Score 1460.5; DB 1; Length 691;
Best Local Similarity 44.3%; Pred. No. 7e-89;
Matches 309; Conservative 140; Mismatches 230; Indels 19; Gaps 11;
QY 42 PEKIRNIGISAHISGKTTLTERLVYTGRIAKKHEKQKGVANVDSMLERORGIT 101
DB 6 PLKVRNIGIAHIDAGKTTISERLFTGSHKIGEY-HDG-ALYDMWQEKRGIT 62
QY 102 IOSAATFTWMDKVNINIIDTPGHVDTIEVERALRVLDGAVLVLCAGVGQCQTIVARQ 161
DB 63 ITSATTCFMKHQNLIDTPGHVDTIEVERAKVLDAVAVFCVSGVGQVQSETVARQ 122
QY 162 MKRIVNPLTFTINKLDRGNSNPARALQMRSKLNNTFMQIPMLGKFNKIVULIEBR 221
DB 123 ANKYGVPRIVFNKMDRIGANFYVEDQIRNRLKANPVLQIPIGAEDNFKGVIDLVTK 182
QY 222 AIFYDPSQIVRY--GEIPAPELRAATDHRQELIECVANSBOLGEMPLEKIPISDL 279
DB 183 ALVWEDD-TKPTDYKEIIPALKEKAEYRTKMLEAVSETSDLMEXYLGSELSLEFI 241
QY 280 KLAIRATLKRSFTFVPLGSAKKNKQVPLDAVLEVLNPSEVQNYALNKKDSKEKT 339
DB 242 KTGIGAGCLSLISIVMLGTAFAKKNQVQPLDAVAVIAPARDEVANI-----KGEYDGT 296
QY 340 KILMNSSRHNSHPVGLAFLPEVGRF-GQLYVRSYQELKKGGDTIYTRTKKVRLORL 398
DB 297 EVSVKST--DDEFAFLAFKIMTDPFVQQLFFVRVYRCLBESGYAYNSTDKKERIGRL 354
QY 399 AMHADMWASTEEVYAGDICALFGI-DCASGDTPTDANSLSMESIHVPDPVISIAWK 457
DB 355 LKMSNKE-EIKVLYAGEIGAVGLKDTLTGTLASEKDK-VILERMDFDPVLSAVE 412
QY 458 PSNKNDLEKFSKIGIRFTREDPTFKYPTDENKEITVIGMGLHLEIYAQRLEREYGCPC 517
DB 413 PTKVADQEMSLALKLKLQEDPSFRVSTDESGQITISGMGLHLEIYDRLRFKVEA 472
QY 518 ITRKRVAFRETTITAPVPPDFTHKKQSGAGAGYGVIGVLEPLDPEDYTKLEFSDETGS 577
DB 473 EVGQPVAVYRETIKVDKVEKHAQSGRGQYGHVFLRLBEPG--SGEYFVNDIKG 530
QY 578 NIPKQVPAVEKGLDACEGKPLSGHLSGLRFLVLDGAAHMDNSNISFIRAGEGALKQ 637
DB 531 VLPKEIIPAVDKGVQDALQNGVLAGYVVEDKAVIYDOSYHEVDSSEMAFKLAASMGFKE 590
QY 638 ALANATLCILEPIMAVEVAPNEFOGVIAGINRRHGVITGQDGVEDYFTLYADVPLNDM 697
DB 591 GARKGAVILEPMWVVEETPEDYMGVMDLNRRIIEGMEETATGKTV 650
QY 698 FGSTELASCTEGKGEYTMESRYOPLPSTQEDVYINK 735
DB 651 FGSTDLRSQOTGRATYSMEFDHYDEKPNVADRIYK 688

RESULT 10
EFG_SPIPL
ID EFG_SPIPL STANDARD; PRT; 697 AA.
AC P13550;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Elongation factor G (EF-G).
GN FUSA OR FUS.
OS Spirulina platensis.
OC Bacteria; Cyanobacteria; Oscillatoriales; Spirulina.
OX NCBI_TaxId=1156;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=89364697; PubMed=2505055;

RA Butarelli F.R., Calogero R.A., Tiboni O., Gualerzi C.O., Pon C.L.;
 RT "Characterization of the str operon genes from *Spirulina platensis*
 and their evolutionary relationship to those of other prokaryotes";
 RL Mol. Gen. Genet. 217:97-104(1989).
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
 OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
 RIBOSOME.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC EF-G/EF-2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL, X15646; CAA3672.1; -.
 CC DR PIR, S04390; S04390.
 CC DR HSSP, P13551; 2EFG.
 CC DR InterPro, IPR000450; EF-G.
 CC DR InterPro, IPR000460; EFG_C.
 CC DR InterPro, IPR004161; EFTU_D2.
 CC DR InterPro, IPR000795; EF_GTPbind.
 CC DR InterPro, IPR005225; Small_GTP.
 CC DR Pfam, PF00009; GTP_EFTU; 1.
 CC DR Pfam, PF00679; EFG_C; 1.
 CC DR Pfam, PF03144; GTP_EFTU_D2; 1.
 CC DR TIGRFAMs, TIGR00231; small_GTP; 1.
 CC DR TIGRFAMs, TIGR00484; EF-G; 1.
 CC DR PROSITE, PS00301; EFATOR GTP; 1.
 CC DR Elongation factor: Proteins biochemistry: GTP-binding.
 CC NP_BIND 17 24 GTP (BY SIMILARITY).
 CC FT NP_BIND 81 85 GTP (BY SIMILARITY).
 CC FT NP_BIND 135 138 GTP (BY SIMILARITY).
 CC SQ SEQUENCE 697 AA; 76775 MW; 28C6DA1837DEB2C CRC64;

Query Match 37.3%; Score 1454; DB 1; Length 697;
 Query Local Similarity 44.0%; Pred. No. 1.9e-88;
 Matches 309; Conservative 138; Mismatches 234; Indels 22; Gaps 11;

QY 41 IPNKKIRNIGISAHNIDSGKTLTERLYTGRYIAKHEVKGQGVANWDSMELEKORCI 100
 DB 5 ILEEVNRNIGIAAHIDAGKTTTTERLTFYSGVHNKGEV--HEGT-AVIDMMAQEREREI 61
 QY 101 TIQSAATFTMMKQVNIINIIDTGHVDFTIEVERALRVNLGAVLVLCAGVCCQMTVNR 160
 DB 62 TITAAAIISTSWLDHRINIIDTGHVDFTIEVERSMKVLGVIAVFCVGVQPOSESTWR 121
 QY 161 QMKRNVNPLFTFINKLDRMGSNPARALQOMRSKLNNTAFMOLPMGLSENFSGYDLLE 220
 DB 122 QMERQVNPRIATINKMDRTGADPEKVGQIRRLRLRANAVPIQVPGRESDFGLVDLVAM 181
 QY 221 RAIVYDGFPSQIVRYG-EIPAEIRAAATDROBELIECVANSDEQLEMLEE---KIPS 275
 DB 182 KTYLVNTDGLDIOVSDEIRPEVQDVAVERKLEAVATEDEALMEKLEQLEGGEALT 241
 QY 276 ISDLKLAIRATLKRSFTVPVGLSALKNGVOPPLDAVLEYLPNBEVQNVAILNKDSS 335
 DB 242 EEEIRHSLKQGTIKGLIVPVGSSSRKNGVORLLDAVDYLPAPTEVP--PIKGLVLPDG 299
 QY 336 KEKTIILNMSRHNHSPFVGLAFLPEVGRGQLTYVRSVQGLKKKGDITYNRTKRYNL 395
 DB 300 EBGVRYADDA-----PLSALAFKWAADPYGRITFRVSVGLQKSYIYATKNNKEET 354
 QY 396 QRLAMHADMWEASTEEVYAGDICALFGI--DCASGDTFTDKANSGLMSIHPVPVLSI 454
 DB 355 SRLIYLVKSD-ERIVVEELRAAGDLGALGKDLTGTLGTDENANS--ILLESIVPREVISV 412
 QY 455 AKKPSKNDLEKPSKIGRFTBEDPFPKYVFTTENKEIYISGMGLHLEIYQRLEREYX 514

DB 413 AVEPKTKODMEKLSKALOSLEEDPTFRVSIIDSETNQTVIAGMGLHLEILVDMLREFK 472
 QY 515 CPCICGKPKVAFRETRITAPVPDFDTHKKQSGAGQYGVIGLBLEPDEYTKLEPSDET 574
 DB 473 VEANGAQAQVAVRETRISITREGFPIRQSGKQYGVNVILEBGEFG--SGEPFVSKI 530
 QY 575 FGSNIKQFVPAVEKFLDACERKGPLSGHKLSGRLVLDQAHMWNDSNISFIRAGECA 634
 DB 531 VGSVPEKVIYINPAQGMKEACESGVIAGYPLIDVATLVDSYHEVDSEMAFKIAGSMA 590
 QY 635 LKQALANNTCLERIMAVEVVAPEFQGVITAGINRHGVITQ--DGVEDFTLYADV 692
 DB 591 IKNGVTKASPVILEPMKVEVEVPDFIGNVIGDLSNRGQIEGQETQSQSIARVAKV 650
 QY 693 PLNDVFGYSTELRSTCEKGEYTMESRYOCLPSTQSDVINK 735
 DB 651 PLATMFGIATDIRSTQGRGVPSMFSHYEEVPSVAETIIAK 693

RESULT 11
 EFG_HAEIN
 ID EFG_HAEIN STANDARD; PRT; 699 AA.
 AC P43525;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Elongation factor G (EF-G).
 GN FUS4 OR PUS OR HT0579.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / K120 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merriek J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirey C.A., Liu L.-I., Glodok A., Kelley J.M.,
 RA Weisman J.F., Phillips C.A., Spriggs T., Hedblom R., Cotton M.D.,
 RA Usterbeck T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of *Haemophilus influenzae*
 Rd".
 RL Science 269:496-512(1995).
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLOCATION
 OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
 RIBOSOME.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC EF-G/EF-2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR PRINTS; PR00315; ELONGATNFACT.
 DR TIGRFAMs; TIGR00231; small GTP; 1.
 DR TRIGRFAMs; TIGR00484; EF-G_1.
 DR PROSITE; PS00301; EFATOR_GTP; 1.
 DR Elongation factor; Protein biosynthesis; GTP-binding;
 KW Complete proteome.
 FT INIT MET 0 BY SIMILARITY.
 FT NP_BIND 16 23 GTP (BY SIMILARITY).
 FT NP_BIND 87 91 GTP (BY SIMILARITY).
 FT NP_BIND 141 144 GTP (BY SIMILARITY).
 SQ SEQUENCE 699 AA; 77132 MW; FBBAD633C0F62801 CRC64;
 Query Match 37.3%; Score 1453; DB 1; Length 699;
 Best Local Similarity 45.1%; Pred. No. 1,9e-88;
 Matches 318; Conservative 124; Mismatches 235; Indels 28; Gaps 13;
 QY 42 PNEKRNIGISAHIDSGKTTLTERTLYTGRARAKHEVKGKDGAVGAVMSMELEORGIT 101
 DB 5 PIERVNNIGISAHIDAGKTTTERTLFTYGVSHKIGEV--HDG-AATMDMEQEGERGIT 61
 QY 102 IGSATFTFMKKDV-----NINIDTGHVDFTEVERALRVLDGAVLTCAGVGVOQ 154
 DB 62 ITSATFTFMKSGSQPPOHRINVIDTGHVDFTEVERSMRVLDGAVVYCAVGVOQ 121
 QY 155 TWTVNRQMKRYNVPPLTINKLDRMGSNPARALQOMRSKLNHTAFMOI PMGLEGNFKGI 214
 DB 122 SEVWRQANKYVPRAPAFVNMKDRGTANFLRVGEQLKTRGANAIPLOQPVGAEMFRTV 181
 QY 215 VDLIERALYFGDGDSQIVRYGEIPAELEAAATDHRQELIECVANSDEQLEGMPLEEK 273
 DB 182 VDLIKKAIWNADQGMFTYEVPANMQADCEENKONLYEAAASAELEMEKYLGGD 241
 QY 274 PSTISDKLIRATLRKSPFVFLGSALKKGVQPLDALTXYLPMPSEVQNAIILKKD 333
 DB 242 LTREERKSAIRQVLANEIIITVCGSAFKKGVQAMLDAAVEYLPAPTDLP--AIKGINP 299
 QY 334 DSKEKTKILMNSRNHS--HPFVGLAFPLEVGRF--GQLTVRSYOGELKKKDTTYNTRR 390
 DB 300 DERE-----GERHNSDESPFSLAFKLTADPFVGNLTFRRYSGVINSGDVLSVNG 352
 QY 391 KATRLRLAMADMEASTEEVYADICALFEI--DCASGDTTDRANSGLSMESHVDP 449
 DB 353 KRERFRIVQMHANKKE--EIKERAGIAAAILKQVLTGDTLC-AIDAPILIERMEFPE 410
 QY 450 PVSIAKPKSNKNDLEFKSGIGRTFRBDPTFKVYFDTEKETEVIKSGMELHETIAQRL 509
 DB 411 PVSISAVEPTKADQEKMGIALGRLAQEDSFVHTDESGEITISGMELHIDITVDRK 470
 QY 510 EREYGCPCITGKPKVAFRETTITAPV--PDPFHKKOSGAGQXGVILGLEPDPEDYTK 568
 DB 471 KRERFVEANIGKQVSRERITRTVADVDEKAKQSGRGQYGHVVIDLYPLDPEG--PGV 529
 QY 569 EREDETFGSGNIPQGFPAVEKGFADACEKPSLGHTLSGLRFLQDGAHMDVDSNEISRT 628
 DB 530 EFNENKGGVTPGEYIPAVDKGQIOELKSGPLAGYVVDIGVLAHFGSYHDVDSSEIAR 589
 QY 629 RAGEGALKOALNAATCILEPIMAVEVQVNEFOGOVIGINRHGVITGODVEVEYFL 688
 DB 590 LNASLAFKAAFSQANPVLLEPIKMEVEYTPPEYVGDVIGDLSRRRAMVNGQEANBEVVKI 649
 QY 689 YADVPLNDMGVSTELRSCTEGKGEYTMESRYQPCLPSTOEDVI 733
 DB 650 YAEVPLSEMGVATDLRSQGRASYSMEPLKAAEAAPTVAALVI 694

DE Elongation factor G 2 (EF-G 2).
 GN FUSB OR FUS2 OR RSP0804 OR RS06124.
 OS Ralstonia solanacearum (Pseudomonas solanacearum).
 OG plasmid megaplasmid.
 OC Bacteri; Proteobacteria; beta subdivision; Ralstonia group;
 OC Ralstonia.
 OX NCBI_Taxid=305;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RX MEDLINE=21681879; PubMed=11823852;
 RA Salanoudat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
 RA Arlat M., Billault A., Broctier P., Camus J.C., Catolico L.,
 RA Chandler M., Cholet N., Claudel-Renard C., Cunnac S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sigmier P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weisenbach J., Boucher C.A.;
 RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
 RL Nature 415:497-502(2002).
 CC -1- FUNCTION: This protein promotes the GTP-dependent translocation of
 the nascent protein chain from the A-site to the P-site of the
 ribosome.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC EF-G/EF-2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AL646081; CAD17955.1; -
 DR InterPro; IPR004540; EF-G.
 DR InterPro; IPR000640; EF-G.
 DR InterPro; IPR004161; EFTU_D2.
 DR InterPro; IPR000795; EF_GTPbind.
 DR InterPro; IPR005225; Small_GTP.
 DR Pfam; PF00679; ERG_C_1.
 DR Pfam; PF00009; GTP_EFTU; 1.
 DR Pfam; PF03144; GTP_EFTU_D2; 1.
 DR PRINTS; PR00315; ELONGATNFACT.
 DR TIGRFAMs; TIGR00484; EF-G_1.
 DR TIGRFAMs; TIGR00231; small GTP; 1.
 DR PROSITE; PS00301; EFATOR_GTP; 1.
 DR Elongation factor; Protein biosynthesis; GTP-binding; Plasmid;
 KW Complete proteome.
 FT NP_BIND 17 24 GTP (BY SIMILARITY).
 FT NP_BIND 88 92 GTP (BY SIMILARITY).
 FT NP_BIND 142 145 GTP (BY SIMILARITY).
 FT NP_BIND 142 145 GTP (BY SIMILARITY).
 SQ SEQUENCE 700 AA; 77619 MW; 53CC4CA877EE919 CRC64;
 Query Match 37.3%; Score 1453; DB 1; Length 700;
 Best Local Similarity 45.2%; Pred. No. 2,2e-88;
 Matches 320; Conservative 126; Mismatches 230; Indels 32; Gaps 14;
 QY 42 PNEKRNIGISAHIDSGKTTLTERTLYTGRARAKHEVKGKDGAVGAVMSMELEORGIT 101
 DB 6 PIERVNNIGISAHIDAGKTTTERTLFTYGVNHKIGEV--HDG-AATMDMEQEGERGIT 62
 QY 102 IGSATFTFMKKDV-----NINIDTGHVDFTEVERALRVLDGAVLTCAGVGVOQ 154
 DB 63 ITSATFTFMKSGSQPPOHRINVIDTGHVDFTEVERSMRVLDGACVYDVSQVGVOQ 122
 QY 155 TWTVNRQMKRYNVPPLTINKLDRMGSNPARALQOMRSKLNHTAFMOI PMGLEGNFKGI 214
 DB 123 SEVWRQANKYVPRAPAFVNMKDRGTANFLRVGEQLKTRGANAIPLOQPVGAEDHFRGV 182
 QY 215 VDLIERALYFGDGDSQIVRY--YGEIPAELEAAATDHRQELIECVANSDEQLEGMPLEEK 272
 DB 183 VDLIKKAIWVD--DASQVKEVYIDIPEELRATQDEWHDKNVEAAAEADALLEKYLGE 241

AC P57593;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Elongation factor G (EF-G).
 GN F05A OR B0527.
 OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 OS symbiotic bacterium).
 OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxId=118099;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Tokyo 1998;
 RX MEDLINE=20445173; PubMed=10993077;
 RA Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 RT "Genome sequence of the endocellular bacterial symbiont of aphids
 Buchnera sp. AS.";
 RL Nature 407:81-86(2000).
 CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT TRANSLLOCATION
 CC OF THE NASCENT PROTEIN CHAIN FROM THE A-SITE TO THE P-SITE OF THE
 CC RIBOSOME.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
 CC EF-G/EF-2 SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: AP001119; BAB13220.1; -.
 DR HSSP: P13551; 1ELO.
 DR InterPro: IPR004540; EF-G.
 DR InterPro: IPR000640; EF-G.
 DR InterPro: IPR004161; EFTU D2.
 DR InterPro: IPR000795; EF GTPbind.
 DR InterPro: IPR005225; Small GTP.
 DR Pfam: PF00009; GTP_EFTU; 1.
 DR Pfam: PF00679; EF-G_C; 1.
 DR Pfam: PF03144; GTP_EFTU_D2; 1.
 DR PRINTS: PR00315; ELONGATINFC.
 DR TIGRPFAM: TIGR00231; small GTP; 1.
 DR TIGRPFAM: TIGR00484; EF-G_1.
 DR PROSITE: PS00301; EFATOR_GTP; 1.
 DR Elongation factor; Protein biosynthesis; GTP-binding;
 KW Complete proteome.
 KM NP_BIND 17 24 GTP (BY SIMILARITY).
 FT NP_BIND 88 92 GTP (BY SIMILARITY).
 FT NP_BIND 142 145 GTP (BY SIMILARITY).
 SO SEQUENCE 702 AA; 78182 MW; F583FEP269ABICF9 CRC64;

Query Match 36.9%; Score 1438; DB 1; Length 702;
 Best Local Similarity 44.5%; Pred. No. 2, 2e-87;
 Matches 317; Conservative 126; Mismatches 241; Indels 28; Gaps 14;

QY 38 SGVIRPEKIRNIGISAHIDSGKTTLIERVLYTGRARAKHEVGRKGVGANDSMLEQ 97
 DB 2 SRTTPSRKRNIGISAHIDAGKTTTERILEYGINHKIGEV--HDG-AATMDMEOE 58
 QY 98 RGTITISAAFTTW-----KDVNIINIDTPGHVDFTIEVERALRVLDGAVLVCAVG 150
 DB 59 RGTITISAAFTTWSCMAKQFKRHRINIITDPGHVFTIEVERSMRVLDGAVVYCAVG 118
 QY 151 VOCQMTVNRQMRKYVVPFLTFINKLDRMGSSPARALQGRSKYLNHTAFMOJPMGLEGN 210
 DB 119 VQPSSEVNRQANKYVVPRIAFVKNQDRMGANFLKVKQIKIRLGANPVLQATGAEPT 178
 QY 211 FKGIVLLIERAYF-DGPEQSQVIRGEIPIAELEAAATDHRQLLIECVANSDEQLCEML 269
 DB 179 FVGVDLICKAKAVHWDSDQGVTFVYNDIPPEMIELSKKNQNLIESAVESNEDLLEKYL 238

QY 270 EEKIPISDLKLAIRATIKRSFTFVPIGSALEKNGKQPLIDAVLEVLPNPEVQNT-AI 328
 DB 229 NODRISEBEIISALAKRALNNEIVILITGSAFKKNGVALLDAILEFIPANDIDINGI 298
 QY 329 LNKKDSKERTKILNNSRRHNSHPVGLAFPLEVGRF-GQLTYVRSYGEELKCDITYNT 387
 DB 229 LN----DVEQTPAIRNSD--DKAPPSALAFIASDPFGNLTFFRVSGVSKGDTVNS 352
 QY 388 RTRKRVRLQRLARMADMMBAESTEVEVAGDICALFGI--DCASGDFFTKANSGLMESIH 446
 DB 353 AKQREERGRIVQMPANKRE-EIKEVYAGDIAAIGLDVTTGDLCD-LNDPILEKME 410
 QY 447 VDDPVIASIAKPSNRNDLEKESKIGREFTRDPFKYVFDTENKETVLSGGELEIAYA 506
 DB 411 PEPVISTISVEKTKVDEKNGKALGRAKEDPSFRVNTDQESNTITISGGELEIILI 470
 QY 507 ORLEREYCCPCTIKPKYAFRETTAPV-PRDFTHKKSGAGQYKXIVLEPLDP--E 563
 DB 471 DMKKEFVDAVNGKQVAVRETTINKVEDIEGKHIKOSGGQYGHVIELFPLQPGE 530
 QY 564 DYTKEFDEFTGSIPIKQFVPAVEKGLDACEKGLSGHLSGLRFLQDGAHMDVSN 623
 DB 531 GYL--FVNDIKGVIPSEYISALDKGIQEQKCGPLAGYFVVDIGVRLYFSGYHVDVS 587
 QY 624 EISFIRAGEGALKQALANATLCILEPINAVEVPAFNEFGOVIAGINRRHGVTQDQVE 683
 DB 588 ELAFLTAASAFAKFKQAKPILIEPIKMEVEETPDYMGVIGDINRRRGIEGMKOLS 647
 QY 684 DYFTLYADVPILDMRGYSTELRSCTEGKGYTMEYSRQPCLPISQOEVIYK 735
 DB 648 ISKINACVPLSEMPGVAIDLRISQOTGRASYSMEPLKYIEAPFNISKDIER 699

RESULT 15
 BEG_SALTY STANDARD; PRT; 703 AA.
 ID_EFG_SALTY
 AC P26229;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Elongation factor G (EF-G).
 GN FUSA OR STM3446 OR STY4352.
 OS Salmonella typhimurium, and
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxId=602, 601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhimurium; STRAIN=LT2;
 RX MEDLINE=93013025; PubMed=1398129;
 RA Johanson H.U., Hughes D.,
 RT "Comparison of the complete sequence of the str operon in Salmonella
 RT typhimurium and Escherichia coli.";
 RL Gene 120:93-98(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Milvanev E.,
 RA Ryan B., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 RT LT2.";
 RL Nature 413:852-856(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S. typhi; STRAIN=CT18;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parhail J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:10:22 / Search time 16.6366 Seconds

(without alignment)
4345.419 Million cell updates/sec

Title: US-09-815-379-6

Perfect score: 3898

Sequence: 1 MRLGMAAVALGRGRAPAS.....INKYLMATGOLPVKKKAKN 752

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	3243	83.2	752	1 S40780	translacion elonga
2	2119.5	54.4	750	2 T21534	hypothetical proce
3	1987	51.0	761	1 S61642	translacion elonga
4	1972	50.6	754	2 D96510	probable mitochond
5	1970	50.5	754	2 P84885	hypothetical prote
6	1903.5	48.8	770	2 T50308	probable translaci
7	1765	45.3	695	2 D71283	translacion elonga
8	1762	45.2	693	2 C70167	translacion elonga
9	1493	38.3	694	1 S04429	translacion elonga
10	1489	38.2	692	2 G71847	translacion elonga
11	1487	38.1	692	2 C64669	translacion elonga
12	1472	37.8	692	2 AC2348	translacion elonga
13	1462	37.5	702	2 AD0025	translacion elonga
14	1460.5	37.5	691	2 H81394	translacion elonga
15	1454	37.3	697	2 S04390	translacion elonga
16	1454	37.3	700	2 P64078	translacion elonga
17	1438	36.9	702	2 D84991	translacion elonga
18	1434	36.8	704	2 JCI424	translacion elonga
19	1434	36.8	704	2 AC1005	translacion elonga
20	1431.5	36.7	691	1 EPTMG	translacion elonga
21	1430	36.7	701	2 E70827	translacion elonga
22	1427	36.6	704	1 EFECG	translacion elonga
23	1427	36.6	704	2 G91152	translacion elonga
24	1427	36.6	704	2 C85998	translacion elonga
25	1426.5	36.6	691	2 S75863	translacion elonga
26	1422	36.5	701	2 H87143	translacion elonga
27	1416.5	36.3	701	2 B82006	translacion elonga
28	1416	36.3	701	2 S31150	translacion elonga
29	1414.5	36.3	694	2 AD3346	protein translatio

30	1413.5	36.3	701	2 C81234	translacion elonga
31	1408.5	36.1	705	2 P82534	translacion elonga
32	1407.5	36.1	695	2 H72242	translacion elonga
33	1406	36.1	702	2 G83386	translacion elonga
34	1401	35.9	687	2 A87286	translacion elonga
35	1399.5	35.9	688	2 S73930	translacion elonga
36	1398.5	35.9	699	2 A70300	translacion elonga
37	1398	35.9	699	2 F97721	translacion elonga
38	1398	35.9	699	2 E97594	translacion elonga
39	1398	35.9	699	2 AC2816	translacion elonga
40	1395	35.8	695	2 S76751	translacion elonga
41	1395	35.8	698	2 C82332	translacion elonga
42	1392.5	35.7	692	2 B69628	translacion elonga
43	1388	35.6	695	2 A81762	translacion elonga
44	1387	35.6	695	2 A81406	translacion elonga
45	1387	35.6	699	2 B71723	translacion elonga

ALIGNMENTS

RESULT 1

S40780

translacion elongation factor EF-G, mitochondrial - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: S40780

R:Barker, C.; Makris, A.; Patriotic, C.; Bear, S.E.; Tichlis, P.N.

Nucleic Acids Res. 21, 2641-2647, 1993

A:Title: Identification of the gene encoding the mitochondrial elongation factor G in ma

A:Reference number: S40780; MUID:93324327; PMID:8332461

A:Accession: S40780

A:Status: nucleic acid sequence not shown

A:Molecule type: mRNA

A:Residues: 1-752 <BAR>

A:Cross-references: EMBL: L14684; NID: G310101; PID: AAA41107.1; PID: G310102

C:Genetics:

A:Map position: 2

C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog

C:Keywords: GTP binding; mitochondrion; nucleotide binding; P-loop; protein biosynthesis

F:48-178/Domain: translation elongation factor Tu homology <Etu>

F:54-61/Region: nucleotide-binding motif A (P-loop)

F:175-178/Region: GTP-binding NKXD motif

F:300-302/Region: GTP-binding SKX/L motif

F:60,61,102,175,176,178,300/Binding site: Mg-GTP (Lys, Thr, Thr, Asn, Lys, Asp, Ser) #st

Query Match	Score	Match	Length DB	ID	Description
Query Match	83.2%	Score 3243;	DB 1;	Length 752;	
Beat Local Similarity	84.1%	Pred. No. 1.2e-203;			
Matches	636;	Conservative	37;	Mismatches	75;
				Indels	8;
				Gaps	5;
Qy	1	MRLGMAAVALGRG---	RAPASLGWOKVNWAKCRSSSGV	PNEKIRNIGSAHDS	57
Db	1	MRLGMAAVALGRG---	RAPASLGWOKVNWAKCRSSSGV	PNEKIRNIGSAHDS	58
Qy	58	GKTLTERVLYYTGRIAMHEVKGDGAVMSDEMELEKRGITIGSAATFTMKDVIN	117		
Db	59	GKTLTERVLYYTGRIAMHEVKGDGAVMSDEMELEKRGITIGSAATFTMKDVIN	118		
Qy	118	IIDTPGHVDFTEVERALRVLDGAVLVCAVGVCQOTMTVNRQKRYNVPFLTFINKD	177		
Db	119	IIDTPGHVDFTEVERALRVLDGAVLVCAVGVCQOTMTVNRQKRYNVPFLTFINKD	178		
Qy	178	RMGSPAPALQOMSKLNHNAFVQIPFGLEGRKGIIDLESPAIFPDGFGQIVRDE	237		
Db	179	RMGSPAPALQOMSKLNHNAFVQIPFGLEGRKGIIDLESPAIFPDGFGQIVRDE	238		
Qy	238	IPAEIRAAATDROELIECVANSDEOLGEMLEEKISISDLKAIIRATIKRSFTPVFL	297		
Db	239	IPADIRAAADHRELEIECVANSDEOLGEMLEEKISISVSLKAIIRATIKRSFTPVFL	298		
Qy	298	GSALKKNCVQPLDAVLEFLPNPSEVQNYAIIINKKDSKEKTKILVNSSRHNSHPVGLA	357		
Db	299	GSALKKNCVQPLDAVLEFLPNPSEVQNYAIIINKKDSKEKTKILVNSSRHNSHPVGLA	357		

QY 358 FPLEVGRFQGLTVYRSYQGELEKKDDTYNTRTKRYLORLARMHADMMSEAEYVYAGD 417
DB 358 FPLEVGRFQGLTVYRSYQGELEKKDDTYNTRTKRYLORLARMHADMMSEAEYVYAGD 416
QY 418 ICALFGIDCASGDTFTDKANSGLSMESIHVPDPVISIANKPSNKDLEKFSKIGCFTR 477
DB 417 ICALFGIDCASGDTFTDKANSGLSMESIHVPDPVISIANKPSNKDLEKFSKIGCFTR 476
QY 478 DPFKRYPTENKETYISGMGELHLEIYQRLEREGCCCTGCKPVARETITAVPPD 537
DB 477 DPFKRYPTENKETYISGMGELHLEIYQRLEREGCCCTGCKPVARETITAVPPD 536
QY 538 FTHKSGGAGA-GQYKGYIVLEPLDEPDYTKLEFSDTFGSNI PKQFVPAVEKFLDACE 596
DB 537 ILHIKSSRRVPAVMGKIVGLEPLADEDPKLEFSETGTGANVPKQFVPAVEKFLDACE 596
QY 597 KGPLSGHKLGLRFLVLDGAAHNVDSNEISFIRAGBALKQALANATLCILEPIMAVEV 656
DB 597 KGPLSGHKLGLRFLVLDGAAHNVDSNEISFIRAGBALKQALANATLCILEPIMAVEV 656
QY 657 APNEFGQVAGINRRHGVITGDDGVEDYFTLYADVPLNDMFGYSTELASCTEGKEEYTM 716
DB 657 APNEFGQVAGINRRHGVITGDDGVEDYFTLYADVPLNDMFGYSTELASCTEGKEEYTM 716
QY 717 EYSRYOPLSTQEDVINKYLEATGOLPYKKGAKN 752
DB 717 EYSRYOPLSTQEDVINKYLEATGOLPYKKGAKN 752

RESULT 2

T21534
hypothetical protein F29C12.4 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 02-Sep-2000
C:Accession: T21534
R:Steward, C.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19436
A:Accession: T21534
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-750 <WIL>
A:Cross-references: EMBL:Z81519; PDB:CA04216.1; GSPDB:GN00020; CESP:F29C12.4
A:Experimental source: clone F29C12
C:Genetics:
A:Gene: CESP:F29C12.4
A:Map position: 2
A:Introns: 16/3; 48/1; 287/3; 438/3; 531/2
A:Superfamily: translation elongation factor G; translation elongation factor Tu homolog
F:45-175/Domain: translation elongation factor Tu homolog <ETU>

Query Match 54.4%; Score 2119.5; DB 2; Length 750;
Best Local Similarity 54.9%; Pred. No. 2.5e-130;
Matches 414; Conservative 125; Mismatches 198; Indels 17; Gaps 6;
QY 9 VALGRGRPASLGWQKQVNMKACRWS-----GVTPNEKIRNIGSAHDSK 59
DB 1 WSLGTFPSAVH--SRQDANVLRFPASNEASVVPVGRTERIKRIGISAHDSK 57
QY 60 TLLTERVLYYTGRIAKMHEVKGQGVAVWDSMELEKRGITIQSAAATPTMMKDVNINI 119
DB 58 TLLTERVLYYAGRIDMHEVKGQGVAVWDSMELEKRGITIQSAAATPTMMKDVNINI 117
QY 120 DTPGVADFTIEVERALRVLDGAVLVICAVGVCCQMTNRRQKRYNVPFLFINKLDM 179
DB 118 DTPGVADFTIEVERALRVLDGAVLVICAVGVCCQMTNRRQKRYNVPFLFINKLDM 177
QY 180 GSNPARALQOMS KILNHTAFMOI PMGLSGNFKGIDLI BEBRIYFDGDPQSQIVRYGEIP 239
DB 178 GATPLKALGLCKLKNHNAALHLPLGKDSNFGIYDLVGHALLYEGSGGLVKKDEIP 237
QY 240 AELRAAATDHROELIECVANSDEQLGEMPLEEKIPISIDLKLAIRPATKRSFTFVLGS 299

DB 228 KDLRYEADDRQGLLEHIANVDETLGEMFLNDQFPNVQIHEALRRYVKAFAFVPLSGS 297
QY 300 ALKXGVPLDVALEIYLPNSEVQVNAIILKKDSDKEKTKILNMSRHNPGLAFLP 359
DB 298 ALKXGVPLDVALEIYLPNSEVQVNAIILKKDSDKEKTKILNMSRHNPGLAFLP 356
QY 360 LEVGRFQGLTVYRSYQGELEKKDDTYNTRTKRYLORLARMHADMMSEAEYVYAGD 418
DB 357 LEVGRFQGLTVYRSYQGELEKKDDTYNTRTKRYLORLARMHADMMSEAEYVYAGD 414
QY 419 ICALFGIDCASGDTFTDKANSGLSMESIHVPDPVISIANKPSNKDLEKFSKIGCFTR 478
DB 415 ICALFGIDCASGDTFTDKANSGLSMESIHVPDPVISIANKPSNKDLEKFSKIGCFTR 474
QY 479 DPFKRYPTENKETYISGMGELHLEIYQRLEREGCCCTGCKPVARETITAVPPD 538
DB 475 DPFKRYPTENKETYISGMGELHLEIYQRLEREGCCCTGCKPVARETITAVPPD 534
QY 539 FTHKSGGAGA-GQYKGYIVLEPLDEPDYTKLEFSDTFGSNI PKQFVPAVEKFLDACE 598
DB 535 FTHKSGGAGA-GQYKGYIVLEPLDEPDYTKLEFSDTFGSNI PKQFVPAVEKFLDACE 594
QY 597 KGPLSGHKLGLRFLVLDGAAHNVDSNEISFIRAGBALKQALANATLCILEPIMAVEV 658
DB 595 KGPLSGHKLGLRFLVLDGAAHNVDSNEISFIRAGBALKQALANATLCILEPIMAVEV 654
QY 657 APNEFGQVAGINRRHGVITGDDGVEDYFTLYADVPLNDMFGYSTELASCTEGKEEYTM 718
DB 655 APNEFGQVAGINRRHGVITGDDGVEDYFTLYADVPLNDMFGYSTELASCTEGKEEYTM 714
QY 717 EYSRYOPLSTQEDVINKYLEATGOLPYKKGAKN 751
DB 715 EYSRYOPLSTQEDVINKYLEATGOLPYKKGAKN 748

RESULT 3

S61642
translation elongation factor EF-G, mitochondrial - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: protein L2195; protein YLR069c
C:Species: *Saccharomyces cerevisiae*
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001
C:Accession: S61642; S20179; S64897; S64901; S17025
R:Ureastarazu, L.A.
submitted to the EMBL Data Library, December 1995
A:Reference number: S61642
A:Accession: S61642
A:Molecule type: DNA
A:Residues: 1-761 <URR>
A:Cross-references: EMBL:X94607; NID:g1181264; PDB:CAA64315.1; PID:g1181269
R:Vanburen, A.; Ackerman, S.H.; Tzagoloff, A.
Eur. J. Biochem. 201, 643-652, 1991
A:Title: Mitochondrial translational initiation and elongation factors in *Saccharomyces*
A:Reference number: S20177; MUID:92037620; PMID:1935960
A:Accession: S20179
A:Molecule type: DNA
A:Residues: 1-65, '1', 67-232, 'V', 234-477, 'S', 479-628, 'P', 630-761 <VAM>
A:Cross-references: EMBL:X58378; NID:g31917; PDB:CAA41267.1; PID:g31918
R:Andre, B.; Ureastarazu, L.A.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64872
A:Accession: S64897
A:Molecule type: DNA
A:Residues: 1-761 <AND>
A:Cross-references: EMBL:Z73241; NID:g1360421; PDB:CAA97626.1; PID:g1360422; GSPDB:GN00
R:Polh, T.M.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64899
A:Accession: S64901
A:Molecule type: DNA
A:Residues: 1-761 <POH>
A:Cross-references: EMBL:Z73241; NID:g1360421; PDB:CAA97626.1; PID:g1360422; GSPDB:GN00

A:Experimental source: strain S288C
 C:Genetics:
 A:Gene: SGD:MEF1; MIPS:YLR069c
 A:Cross-references: SGD:S0004059; MIPS:YLR069c
 A:Map position: 12R
 C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog
 C:Keywords: GTP binding; mitochondrion; nucleotide binding; P-loop; protein biosynthesis
 P:71-805/Domain: translation elongation factor Tu homology <ETU>
 P:177-84/Region: nucleotide-binding motif A (P-loop)
 P:202-205/Region: GTP-binding NKD motif
 P:327-329/Region: GTP-binding SAK/L motif

Query Match 51.0%; Score 1987; DB 1; Length 761;
 Best Local Similarity 56.2%; Pred. No. 1.1e-121;
 Matches 392; Conservative 117; Mismatches 172; Indels 16; Gaps 8;

Qy 45 KIRNIGISAHIDSGKTTTLTERVLYYGRIRAKHEVKGKGVGAVNDSMELERQGITIOS 104
 Db 69 KIRNIGISAHIDSGKTTTLTERVLYYGRIRAKHEVKGKGVGAVNDSMELERQGITIOS 128
 Qy 105 AATFTMW---KDVNINIDTGFHVDFTIEVERALRVLDGAVLVCAVGVCQMTTVNR 160
 Db 129 AATYGSMDKGNHNLIDTGFHVDFTIEVERALRVLDGAVLVCAVGVCQMTTVNR 188
 Qy 161 QMKRYNVPFLTFINKLDRMGSNPARALQOMSKLNHTAFMOJPMGLEGNFGYVDLIEE 220
 Db 189 QMKRYNVPFLTFINKLDRMGSNPARALQOMSKLNHTAFMOJPMGLEGNFGYVDLIEE 248
 Qy 221 RAIYDGDGDSQIVRGEIYPAELRAAATDRQELIECVANSDEQLEMFLFEKIPISDLK 280
 Db 249 VAIYNGDGEIIEKGPVENLKLPMLEKROLLIEFLADVDEMEFLFEKIPISDLK 308
 Qy 281 LAIRBATLKRSFTVPFLGSLKNKGVOPLDAVLEYLPNPSVQVYALINKDDSEKTK 340
 Db 309 DAIRSTIRASTFTVPYMGSLANTGIGPLDAIVDLPNPSVQVYALINKDDSEKTK 364
 Qy 341 IIMNSSRNHSHPFVGLAFPLEVGRFQQLTYVNSYOGELKKGDTIYNTTRTKKVRQLRLAR 400
 Db 365 V--NLVPAVQGFVGLAFPLEVGRFQQLTYVNSYOGELKKGDTIYNTTRTKKVRQLRLAR 422
 Qy 401 MHAADMEASTEEVYAGDICALFGIDCASGDTFTDKANSLSMESIHVPVPVSIAMKPSN 460
 Db 423 MHSSEME-DVDEVGSEIATFGIDCASGDTFTD--GSVQYSMSMVPVPAVVSLSITPNS 480
 Qy 461 KNDLEKFSKIGRFPREDPTFKYVPFTEKETEVIISGMELHLIYAQRLEREYGCPTIG 520
 Db 481 K-DASNFSKALNRFQKEDPTFKYVPFTEKETEVIISGMELHLIYAQRLEREYGCPTIG 539
 Qy 521 KEKVAFRETIITAPVPEDFTHKQSGAGQYGVIVGLLEPLDEPDYTKLE-PSDETFGSNI 579
 Db 540 KQVQVRESITIPADPDYTHKKQSGAGQYGVIVGLLEPLDEPDYTKLE-PSDETFGSNI 597
 Qy 580 PKQFPYAVKGLDACEKGPLSGHKLISGLRFLVLDGAAHMDNSNISFTIRAGEGALKQAL 639
 Db 598 PKQYLAACKGFEVECEKGPLIGHRLVDYMLINGAHIVDSNELSFTIATMSAFRDLF 657
 Qy 640 ANATCILPEPIAIVEVVAENRFGQVYIAGINRHHGVIYTCQGVEDYFTLYADVPILNDMG 699
 Db 658 LRAQVYINRPIINNVSTYSPNEFGVNIIGLAKLQAVITQCTENGHDEFTLAKCALSTWFG 717
 Qy 700 YSTELRSCTEGKGEYTMESRYOPLPSTOEDVINKY 736
 Db 718 PATSLRASTQKGFEFLFSHYAFAPAHQKELISEF 754

RESULT 4

D96510
 probable mitochondrial elongation factor [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
 C:Accession: D96510
 R:Riflogies, A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chlin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, B.; Kim, C.
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucos, J.S.; Malt, R.; Marzilli,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Saldeeb, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: D96510
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-754 <STO>
 A:Cross-references: GB:A8005173; NID:g11321780; PIDN:AAG34257.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog
 C:Keywords: mitochondrion

Query Match 50.6%; Score 1972; DB 2; Length 754;
 Best Local Similarity 54.8%; Pred. No. 1.1e-120;
 Matches 385; Conservative 118; Mismatches 187; Indels 12; Gaps 5;

Qy 35 WSSGVIPNEKIRNIGISAHIDSGKTTTLTERVLYYGRIRAKHEVKGKGVGAVNDSMEL 94
 Db 58 WKES---MDKIRNIGISAHIDSGKTTTLTERVLYYGRIRAKHEVKGKGVGAVNDSMEL 113
 Qy 95 ERQGITIOSAATFTPMKDVNINIIDTGFHVDFTIEVERALRVLDGAVLVCAVGVCQ 154
 Db 114 ERQGITIOSAATFTPMKDVNINIIDTGFHVDFTIEVERALRVLDGAVLVCAVGVCQ 173
 Qy 155 TMTVROKRYNVPFLTFINKLDRMGSNPARALQOMSKLNHTAFMOJPMGLEGNFGI 214
 Db 174 SITDRQRRKEVERVAINTLDRMGADPMKYLQAKKHHSAVQVPIGLENPQGL 233
 Qy 215 VDLIERAIYDGDGDSQIVRGEIYPAELRAAATDRQELIECVANSDEQLEMFLFEKIP 274
 Db 234 IDLIHVKAYPFHSGSGEVNVAAGDIPADMEGLVAEKRELIETVSEVDVDAEKFLNDEPV 293
 Qy 275 SISDLKLAIRBATLKRSFTVPFLGSLKNKGVOPLDAVLEYLPNPSVQVYALINKDD 334
 Db 294 SASLEBAIRATTAQGTVPVFGSAFQNKGVOPLLDVSFLPSPNVNNYAL---DO 349
 Qy 335 SKEKTIIMNSSRNHSHPFVGLAFPLEVGRFQQLTYVNSYOGELKKGDTIYNTTRTKKVR 394
 Db 350 NNNEERVLTSSPDG--PVALARLEBGRGQQLTYLRVYGVIVKKGDPITINANTGKIK 407
 Qy 395 LQRLARHADMEASTEEVYAGDICALFGIDCASGDTFTDKANSLSMESIHVPDPVISI 454
 Db 408 VPLRVHMSNDME-DIGAHAGQIVAVGICASGDTFTD--GSVXYTWSMVPVPAVVSLSITPNS 465
 Qy 455 AMKPSNKNLDEKFSKIGRFPREDPTFKYVPFTEKETEVIISGMELHLIYAQRLEREY 514
 Db 466 AVQVPSKDSGQFISALNRFQKEDPTFKYVPFTEKETEVIISGMELHLIYAQRLEREY 525
 Qy 515 CPCTIGKPKVAFRETIITAPVPEDFTHKQSGAGQYGVIVGLLEPLDEPDYTKLEFSDET 574
 Db 526 VDATGKRNVAFRETIITAPVPEDFTHKQSGAGQYGVIVGLLEPLDEPDYTKLEFSDET 585
 Qy 575 FGSNIPKQFPYAVKGLDACEKGPLSGHKLISGLRFLVLDGAAHMDNSNISFTIRAGEG 634
 Db 586 VQQAIPSGFPIAIEKGFEEANSGSLIGHPVENRIVLTDASHAVDSSELAFKAAIYA 645
 Qy 635 LKQALANATCILPEPIAIVEVVAENRFGQVYIAGINRHHGVIYTCQGVEDYFTLYADVP 694
 Db 646 FRLCYTAPRVILPEPVMVLEKVPTEFGVYAGDINRKGKGIYVNDGDDSVITANVPL 705
 Qy 695 NDMEGYSTELRSCTEGKGEYTMESRYOPLPSTOEDVINKY 736
 Db 706 NNMEGYSTELRSCTEGKGEYTMESRYOPLPSTOEDVINKY 747

RESULT 5

P84885
 hypothetical protein At2g45030 (imported) - Arabidopsis thaliana
 C/Species: Arabidopsis thaliana (mouse-ear cress)
 C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
 C/Accession: P84885
 R./Lin, X.; Kall, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanakken, S.E.; Umayam, L.; Tallon, L.
 eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A./Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A./Reference number: A84420; MIMD:20083487; PMID:10617197
 A/Accession: P84885
 A/Status: preliminary
 A./Molecule type: DNA
 A./Residues: 1-754 <STO>
 A./Cross-references: GB:AE002093; NID:94895248; PIDN:AMD32833.1; GSPDB:GN00139
 C/Genetics:
 A./Gene: At2g45030
 A./Map position: 2
 C./Superfamily: translation elongation factor G; translation elongation factor Tu homolog

Query Match 50.5%; Score 1970; DB 2; Length 754;
 Best Local Similarity 54.8%; Pred. NO. 1.4e-120; Indels 12; Gaps 5;
 Matches 385; Conservative 117; Mismatches 188;
 35 WSSSGVPIPEKINIGISAHIDSGKTLTERVLYTGRIAKMEVKKGKGVGAMDSMEL 94
 58 WKES---MDKLNIGISAHIDSGKTLTERVLYTGRIRHIEHVGROGVGAKMDSMDL 113
 95 ERKRGITIGSAAFTMMKQVNIITPGHVDFTIEVERALRLDGAIVLCAVGVQCC 154
 114 ERKRGITIGSAAFTMMKQVNIITPGHVDFTIEVERALRLDGAIVLCAVGVQCC 173
 155 TMVNVOMKRYNVPFLTFIKLDRMGSPARALQKRSKLNHTAPMOI PMLEGKPKGI 214
 174 STIVDOMKRYNVPFLTFIKLDRMGSPARALQKRSKLNHTAPMOI PMLEGKPKGI 233
 215 VDLIERAIFYDDPSQIVRYGELPAELRAAATDHRQELIECVANSDEQSGEMFLEKIP 274
 234 IDLIHKAIFYDDPSQIVRYGELPAELRAAATDHRQELIECVANSDEQSGEMFLEKIP 293
 275 STIDELKAIIRATLKRSTFTVFGSLAKKQVPLLDVALEIYLPNPEVONTAII LKCKD 334
 294 SALEELAIRATLKRSTFTVFGSLAKKQVPLLDVALEIYLPNPEVONTAII LKCKD 349
 335 SKETKILMNSSHNSHPFVGLAFPLEVGRFGQITVRSYQELTKKGDITYNTRTKKVR 394
 350 NNNEKRYTLTGSPDG--PLVALAFKLEBGRFGQITVRSYQELTKKGDITYNTRTKKVR 407
 395 LQRLAEMADMEASTEEVYAGDICALFGIDCASGDTFTDKANSGLSMESIHVPDVISI 454
 408 VPLVAMHNSDME-DIOEAHAGQIVAVFGIECASGDTFTD--GSVKYTMNSMNPBEVMSL 465
 455 AMKPSKNDLEKSKIGRTREDPTFKYFTDTENKETYISGMGLHLIYAOQLREVG 514
 466 AVQPVSKDSGGPSKALNRFQKEDPTFRVGLDPESSQITIIISGMGLHLIYVERMRERK 525
 515 CPCITGPKVAFRETTAPVPFDTKHKQSGAGQYGVYGLVEPLDPEDTYTLKESDSE 574
 526 VDATVGRKPNRETTTQRAEFYLLHKQSGAGQYGVYGLVEPLDPEDTYTLKESDSE 585
 575 FGSNIPKOPVAVKEGFLDACERKPLSGHKLISGLRVLQDGAHMDVDSNEISFIRAGEA 634
 586 VGGATISGFIPALIEKFKKAAKNSGLIGHVENVRLVLDGASHAVDSSSLAKMAIYA 645
 635 LKQALANATLCLIEPTMAVEVAVAPNEQGVYAGINRRHGVITGDQVEDYFTLVADVL 694
 646 FRLCYTAAPVILIEPTMAVEVAVAPNEQGVYAGINRRHGVITGDQVEDYFTLVADVL 705
 695 NDMFGSTELRSCTEGKGEYMEYSRYQCLPSTQSDVINKY 736
 706 NMFGISTLSRMTQSGEFTMEYKSHAVSNVEVQALVNAY 747

RESULT 6
 T50308
 probable translation elongation factor EF-G SPBC1306.01c precursor, mitochondrial [setml]
 C/Species: Schizosaccharomyces pombe
 C/Date: 09-Jun-2000 #sequence_revision 03-Nov-2000 #text_change 03-Nov-2000
 C/Accession: T50308; T40449
 R./Brown, S.; Harrie, D.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, November 1999
 A./Reference number: Z25060
 A/Accession: T50308
 A/Status: translated from GB/EMBL/DBJ
 A./Molecule type: DNA
 A./Residues: 1-252

 A./Cross-references: EMBL:AL13303; PIDN:CAB62091.1; GSPDB:GN00067; SPDB:SPBC1306.01c
 A./Experimental source: strain 972h(-); cosmid c1306
 R./Lyne, M.H.; Rajandream, M.A.; Barrell, B.G.; Chillingworth, T.; Churcher, C.M.
 submitted to the EMBL Data Library, August 1999
 A./Reference number: Z21929
 A/Accession: T40449
 A/Status: translated from GB/EMBL/DBJ
 A./Molecule type: DNA
 A./Residues: 213-770 <LN>
 A./Cross-references: EMBL:AL109822; PIDN:CAB52624.1; GSPDB:GN00067; SPDB:SPBC409.22c
 A./Experimental source: strain 972h-; cosmid c409
 C/Genetics:
 A./Gene: SPDB:SPBC1306.01c; SPDB:SPBC409.22c
 A./Map position: 2
 A./Genome: nuclear
 C./Superfamily: translation elongation factor G; translation elongation factor Tu homolog
 C./Keywords: mitochondrion
 F/60-208/Domain: translation elongation factor Tu homology <ETU>

Query Match 48.8%; Score 1903.5; DB 2; Length 770;
 Best Local Similarity 54.4%; Pred. NO. 3.2e-116; Indels 33; Gaps 10;
 Matches 390; Conservative 111; Mismatches 183;
 44 EKIRNIGISAHIDSGKTLTERVLYTGRIAKMEVKKGKGVGAMDSMELERKIGITIQ 103
 57 KQIRNIGISAHIDSGKTLTERVLYTGRIRKDHVRKQNVGAMDSMELERKIGITIQ 116
 104 SAATTTM-----KQVNIITPGHVDFTIEVERALRLDGAIVL 145
 117 SAATCTCETERTVDQIEANEKQTPKESKYNINIITPGHVDFTIEVERALRLDGAIVL 176
 146 CAVGVQCCQTMVNRKQRYNVPFLTFIKLDRMGSPARALQKRSKLNHTAPMOI PM 205
 177 CAVGVQCCQTMVNRKQRYNVPFLTFIKLDRMGSPARALQKRSKLNHTAPMOI PM 236
 206 GLEGNFKGIVDLIERAIFYDDPSQIVRYGELPAELRAAATDHRQELIECVANSDEQ 264
 237 GQEDLKEGVNDLQIRALYNNKSGKEKIEISQVPENLIELAKERSALIELADLDEI 296
 265 GEMPLEEKIPISIDKAIIRATLKRSTFTVFGSLAKKQVPLLDVALEIYLPNPEVQ 324
 297 ADIYMEEDPTPEQIMGAIRRTTARKETPVLMGSLANVQSVLDVAVCYLPPNSPVE 356
 335 NVAILNKQDSEKTKILMNSSHNSHPFVGLAFPLEVGRFGQITVRSYQELTKKGDIT 384
 357 NIA-LNAAD--SEKVSILVPS--EKLVALAFKLEBGRFGQITVRSYQELTKKGDIT 410
 385 YNTRTKRYRLQRLAEMADMEASTEEVYAGDICALFGIDCASGDTFTDKANSGLSMES 444
 411 YVNSTTKIKYSLRVAMHNSDME-BIEKVEAGGICALFGIDCASGDTFTDSSVS-YMTS 468
 445 IHVPDVISIAMKPSKNDLEKSKIGRTREDPTFKYFTDTENKETYISGMGLHLI 504
 469 MEVPSPVLSLIPKSK-DTTSFSKALNRFQKEDPTFRVGLDPESSQITIIISGMGLHL 527
 505 YAOQLREYGCPCITGPKVAFRETTAPVPFDTKHKQSGAGQYGVYGLVEPLDPE 563
 528 YVERRREYKQDCETGERVAFRETTAPVPFDTKHKQSGAGQYGVYGLVEPLDPE 587

Qy 564 D----YTKLESDDETFGSGNIPKQVPAVEKGLDACEKPLSGHKLSGRLFLQDGAHHM 619
Db 588 DESGNVVDCEFIINKYTGCTVPTQYIPACKAKAFYEALKKGFLIGHPIKNGRFLVEDGAAYH 647
Qy 620 VDSNEISIFRAGEGALKQALANATLCIEPIMAVEVNAENRQGOVIAGINRRHVITQ 679
Db 648 VDSSEIAFLATITISAFRTAFLOANPWVEPIINNVSTAVEHQGGVIGLNDKRAKTIVDS 707
Qy 680 DGEVDFYLYADVPPLNDMFGSTELRSCTEGKEGYMEYERQPCLPSTQEDVINKY 736
Db 708 DTDEDEFTLQAEVPLANSMSYSDDIRALTKGGEFSEMEFLKYLPAKYQKELVAY 764

RESULT 7
D71283
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 02-Feb-2001
C:Accession: D71283
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
Reon, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770; PMID:9665876
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-695 <Col>
A:Cross-references: GB:AE001248; GB:AE000520; NID:g3323074; PIDN:AAC65735.1; PID:g332307
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0767
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog
C:Keywords: GTP binding; nucleotide binding; P-loop
F:9-139/Domain: translation elongation factor Tu homolog <ETU>
F:15-22/Region: nucleotide-binding motif A (P-loop)
F:136-139/Region: GTP-binding NKXD motif

Query Match 45.3%; Score 1765; DB 2; Length 695;
Local Similarity 50.0%; Pred. No. 3e-107;
Matches 347; Conservative 131; Mismatches 200; Indels 16; Gaps 7;

Qy 47 RNIGISAHIDSGKTLTTERLVYTGRIAMHEVKGKGVAGVAVDSMELERORGITQSA 106
Db 9 RNIGISAHIDSGKTLTTERLVYTGRIAMHEVKGKGVAGVAVDSMELERORGITQSA 68
Qy 107 TETMKADVNIINIDFGHVDFTIEVERALRVLDGAVLVCAVGVQOCTMTVNRQMKRY 166
Db 69 TGVWKGHTINVIDFGHVDFTIEVERSLRVLDGAVLVCAVGVQOCTMTVNRQMKRY 128
Qy 167 VEPFLPINKLDMGSPARALQOMRSLNHNTPAQIPKGLGNFGYIDLLEERAIYD 226
Db 129 VRISEINCDRTGAPFPFCAQLREKSLNHLMLQPLGLDRLEGVIDLILKALYE 188
Qy 227 GDSQIVRYGEIPEALRAATDHRQELIECVANSDEQGLMEFLSEKIPISD--LKLAI 284
Db 189 GESSGAHVRAPIPEQVQAVVKKRDELIAASISFSELNAYIE--GTEQQILRAVR 245
Qy 285 RATLKRSFTVPLGSLAKNKGVOPLLDVAVLEYPNPESEVQVYAIINKKDSKEKTKILM 344
Db 246 AGTIAKEFVFCGSAVKKGIQPLDATTYLLPDETVNTAL--DLDRABEPTL--- 300
Qy 345 SSRNHSHPVGLAFPLEVERFGOLTYVRSYQELKKGGDTTYNTRTKKRLQLRLAMAD 404
Db 301 -STDDAPVVALGFKLEDKGYQLTYVRYQGITKKGAH;FNVRKAKKRVKRLVLRMSN 359
Qy 405 NMEEASTEEYAGDICALFGIDCASGDTFTDKANSGLSMESIHVPDVISIAMKPSKNDL 464
Db 360 QMEDISEGR-PGDYALFVDCASGDTFC--SGDLNATMSMVFVEVISLSTTPKDKSA 417
Qy 465 EKPSKIGRFTREDPTFKVYPTENKETVISGMGLHLEIYAOQLREYGCPCITGPKPY 524

Db 418 DQVSKALNRFKEDPTERSFVDPESNOTITQGMGLHLVDVIERMRREYKCEVETGMPQV 477
Qy 525 AFRETTAPVPEDFTHKKQSGAGQYGVIGVLEPLDEPDYTKLESPDETGSNI PKQFV 584
Db 478 AYREIASRADPNYTHKKQTSQSGQFGVAGFIERIAGQV---EFDQIKGVYIPNEFI 534
Qy 585 PAVEKGLDACEKPLSGHKLSGLRFLVQDGAHHMVDSENEISIFRAGEGALKQALANATL 644
Db 535 PSCDGFRTAVKKGTLIGFPIVGVRTINDGQSHPVDSDMAFOAAIAGFREAVNGAKP 594
Qy 645 CILPEPIMAVEVNAENRQGOVIAGINRRHVITQODGVEDYFTLYAVPLNDMGYSTEL 704
Db 595 VLEPIMKSVSGQEPFGSVFGLINQRGVAVSSADDEQFSRVDAEVLSEMFSTVL 654
Qy 705 RSCTEKGEYMEYERQPCLPSTQEDVINKY 738
Db 655 RSTQGRKREYMEPRAKKAPQGVYDLSIKETOE 688

RESULT 8
C70167
translation elongation factor G (fus-1) homolog -- Lyme disease spirochete
C:Species: Borrelia burgdorferi (Lyme disease spirochete)
C>Date: 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 02-Feb-2001
C:Accession: C70167
R:Fraser, C.M.; Caejens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Iachigra, R.; White
son, D.; Peterson, J.; Kurlavag, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt,
Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.
Nature 380, 580-586, 1997
A:Authors: Smith, H.O.; Venter, J.C.
A:Title: Genomic sequence of a Lyme disease spirochete, Borrelia burgdorferi.
A:Reference number: A70100; MUID:98055943; PMID:9403685
A:Accession: C70167
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-693 <KLE>
A:Cross-references: GB:AE001155; GB:AE000783; NID:g2688448; PIDN:AAC66897.1; PID:g268844
A:Experimental source: strain B31
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog
C:Keywords: GTP binding; nucleotide binding; P-loop
F:7-137/Domain: translation elongation factor Tu homolog <ETU>
F:13-20/Region: nucleotide-binding motif A (P-loop)
F:134-137/Region: GTP-binding NKXD motif

Query Match 45.2%; Score 1762; DB 2; Length 693;
Local Similarity 51.8%; Pred. No. 4.7e-107;
Matches 361; Conservative 126; Mismatches 188; Indels 22; Gaps 10;

Qy 45 KIRNIGISAHIDSGKTLTTERLVYTGRIAMHEVKGKGVAGVAVDSMELERORGITQ 104
Db 5 KIRNIGISAHIDSGKTLTTERLVYTGRIAMHEVKGKGVAGVAVDSMELERORGITQ 64
Qy 105 AATFTMKDVNIINIDFGHVDFTIEVERALRVLDGAVLVCAVGVQOCTMTVNRQMKR 164
Db 65 AATHEWMDPFINIIDFGHVDFTIEVERSLRVLDGAVLVCAVGVQOCTMTVNRQMKR 124
Qy 165 YNVPFLTINKLDMGSPARALQOMRSLNHNTPAQIPKGLGNFGYIDLLEERAIY 224
Db 125 YSVPFLAVNCKDRTGAPFPFCAQLREKSLNHLMLQPLGLDRLEGVIDLILKALYE 184
Qy 225 FDG-DPSQIVRYGEIPEALRAATDHRQELIECVANSDEQGLMEFLSEKIPISDILKAI 283
Db 185 FEGDGTITIE-KETPSDLLEAKSKREIMDLADPDELMEHMEKQVPTETIYAT 243
Qy 284 RRAITKRSFTVPLGSLAKNKGVOPLLDVAVLEYPNPESEVQVYAIINKKDSKEKTKILM 343
Db 244 RTGTALATLCVFFGSAVKKGVOLLDVAVRFLPSPHIDIKTALDINNKEKEIDLK-- 301
Qy 344 NSSRNHSHPVGLAFPLEVERFGOLTYVRSYQELKKGGDTTYNTRTKKRLQLRLAMAD 403
Db 302 ---DNEPLTVALAFKLEDQYQGLTYVRYQGITKKQELINSRTEKFFVGRILIRHA 357
Qy 404 DMMEASTEEY---AGDICALFGIDCASGDTFTDKANSGLSMESIHVPDVISIAMKPSN 460

Db 358 N---:MTEDIEFGSGDIYALFGEICASGDTFCD-PSINYSMTSMEIPPIVSLISVYKPD 412
461 KNDLEKFSKIGRFTREDDPTEKYPFTENKETVISCAGELHLEIYQORLEREGCCICNG 520
413 KKSADMMAKLGFRTEDPTEKTYVIESNETIIOGGBLHLEIYERKREKAEVEFTG 472
Qy 521 KPKVARETITTAVPDPFTHKKQSGAGQYKVIYGLBPLDPEDYTKLEFSDETPGSNIP 580
473 MPQVARETITRAKEFNHYTHKKQSGAGQYKVGAMEPLDKKGET-YEFVNLKKGCVIP 531
Db 581 KQVPAVEKGFLDACEKGPLSGHKSGLRVLQDGAHMDWSNEIFIRAGBALKQALA 640
532 TEYIPSCDKGFQKAMERGTILGPYVDIKTINDGQYHIVDSSDIAFOLAIGAFAREAYE 591
Qy 641 NATLCILEPIMAVEVVAAPNEFOGOVYAGINRHHGVITG--QDGVEDYFTLYADVPINDMF 698
592 KAKPTLLEPIMKTYLLEGPTFEFGNMGLNQRGIIITGSLDGS--SFSKVEAEVPLESMF 649
Db 699 GYSTELRSCTEGKGYTMEYSRYOPLPSTQEDVINK 735
650 GFSTVLRSSITQKAEFSMEFLRYGK-VPSITIFDELNK 685

RESULT 9

S04429
translation elongation factor EF-G - *Synechococcus* sp. (strain PCC 6301)
C:Species: *Synechococcus* sp.
C>Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 19-Jan-2001
C:Accession: S04429
R:Memg, B.Y.; Shinozaki, K.; Sugitara, M.
Mol. Gen. Genet. 216, 25-30, 1989
A>Title: Genes for the ribosomal proteins S12 and S7 and elongation factors EF-G and EF-A:Reference number: S04426; MUID:89281486; PMID:2499762
A:Accession: S04429
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-694 <MEN>
A:Cross-references: GB:X17442; NID:g11204; PIDN:CA43495.1; PID:g1405432
A>Note: the source is designated as *Anacystis nidulans*
C:Genetics:
A:Gene: fus
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog
C:Keywords: GTP binding; nucleotide binding; P-loop; protein biosynthesis
F:11-138/Domain: translation elongation factor Tu homology <ETU>
F:17-24/Region: nucleotide-binding motif A (P-loop)
F:135-138/Region: GTP-binding NKXD motif
F:111-313/Region: GTP-binding SAK/L motif
F:23,24,62,135,138,311/Binding site: Mg-GTP (lys, thr, thr, Asn, lys, Asp, Ser) #sta

Query Match 38.3%; Score 1493; DB 1; Length 694;
Best Local Similarity 44.5%; Pred. No. 1.7e-89;
Matches 315; Conservative 128; Mismatches 229; Indels 36; Gaps 9;
Qy 41 IPNEKIRNIGISAHIDSGLTTLTERVLYTGRIRAKHEVKGKDGAVAMDSELEFORGI 100
5 VPLEKVRNIGIAHIDAGKTTTERTILFYSGVHKIGEY--HDG--NAATDMMEOEKERGI 61
Db 101 TIOSAATPFMMKQVNIITDPGHVDFTEVERALVLDGAVLVCAVGVQCQMTNVR 160
62 TITAAIISTSWKQYRNIITDPGHVDFTEVERSKMVLGAVAVFCSVGQVQOSTVWR 121
Qy 161 QMKRVVPLTFINKLDRMGSNPARALQOMRSKLNHTAFMOIPMGLGKFGKIVLIEER 220
122 QADRVSVPRIVFNKMDRIGADDFKYYGQIRDRVRAANVPICPIABESDPQIYDLVEM 181
Db 221 RAIFPDGDSQIVRYGEIPAELEAAATDHRQELIECVANSDEQGMFLBEPKIPISIDL 280
182 KAHITVNDLGTDLVIDIPAELOETAABWRSKVAEVAETDEALDKYFEDGDLSTEDIX 241
Qy 281 LAIRRTTLKRS---FPFVPLGSLKRGVQPLLDATLEVLNPSVQ-----NVAI 328
Db 242 AGURKVVLLIQGNDRLVPMCGSAFNKGVQLLDVAVELLPSPDIPPIQGTLPDGEVAL 301

Qy 329 LNKDQSKKTKILMNSSRHSNHPFVGLAFPLEVGRFGQLTYYRSYOGELKKGDTIYNTR 388
302 RPSDEA-----PESALAFKIMADPYGRILFVRYSGIILQKGSVYNAT 345
Db 389 TRKRYRLQRLAMADMMEAESTEVEYADICLFSI--DCASGDTFTDKANGLSMESIHV 447
346 KKKKRVSRLLIILKAD--DRIEVDLRAQDLGAVGLKDTFTGDTLDDQNP--IILSELI 403
Qy 448 PDPVLSIMKPSKNDLEKFSKIGRFTREDPTEKYPFTENKETVISCAGELHLEIYQAQ 507
404 PEPVLSAVERPTKNDMEKLSKALALSEBDPTFVSVSDSETNOYVIMGBELHLEILD 463
Db 508 RLREYGCPCITGKPKVAFRETTTAPVPDPFTHKKQSGAGQYKVIYGLBPLDPEDYTK 567
464 RMLREYKVEANIGAQVAVRETIVRAKVAEGKFNQSGKGQYGHVIVLEBPARG--TG 521
Qy 568 LEFSDETGSGNIPKQFPAVEKGFLDACEKGPLSGHKSGLRVLQDGAHMDWSNEISF 627
522 FEFVSKIVGTVPKXYVGPABQGMKETSGLVAGYPLIDIKATLVDSYHDVDSSEMAF 581
Qy 628 IRAGEGALKQALANATLCILEPIMAVEVVAAPNEFOGOVYAGINRHHGVITGQDVEDYFT 687
582 KTAGSMALKEAVRKADPVLEBPMKVEVEVEDPFGSMGVLISRGQIEQATTINGAT 641
Qy 688 LYADVPLNDMPGYSTELRSCTEGKGYTMEYSRYOPLPSTQEDVINK 735
642 VSAKVPFLAMEFYATDLDSMTQGRGIFMERSQYVEVPRVNAETIIAK 689

RESULT 10

G71847
translation elongation factor EF-G (ef-g) - *Helicobacter pylori* (strain J99)
C:Species: *Helicobacter pylori*
A:Variety: strain J99
C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 02-Feb-2001
C:Accession: G71847
R:Alim, R.A.; Lång, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.; Ives, C.; Gibson, R.; Merberg, D.; Miller, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.; Nature 397, 176-180, 1999
A>Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
F:11-138/Domain: translation elongation factor Tu homology <ETU>
F:17-24/Region: nucleotide-binding motif A (P-loop)
F:135-138/Region: GTP-binding NKXD motif

Query Match 38.2%; Score 1489; DB 2; Length 692;
Best Local Similarity 45.7%; Pred. No. 3.1e-89;
Matches 319; Conservative 128; Mismatches 233; Indels 18; Gaps 11;
Qy 42 IPNEKIRNIGISAHIDSGLTTLTERVLYTGRIRAKHEVKGKDGAVAMDSELEFORGI 101
6 PLNIRNIGIAHIDAGKTTTERTILFYGVSHKIGEY--HDG--AATDMMEOEKERGIT 62
Db 102 IOSAATPFMMKQVNIITDPGHVDFTEVERALVLDGAVLVCAVGVQCQMTNVRQ 161
63 ITSAAITCFMDHQINLIDTPGHVDFTEVERSKMVLGAVSVFCSVGQVQOSTVWRQ 122
Qy 162 MKRVVPLTFINKLDRMGSNPARALQOMRSKLNHTAFMOIPMGLGKFGKIVLIEER 221
123 ANKIVSVPRIVFNKMDRIGANFVYVENQIQRLKRNPPPINIPICAEETFTGIVDLVGMK 182
Qy 222 AIYFDGDSQIVRYG--EIPAELEAAATDHRQELIECVANSDEQGMFLBEPKIPISIDL 279

Db 183 ALVWNNK-TWGAAYDVEEIPSDLEKAKQREVLVAVAHFOEALMERTLGGELDIEI 241
Qy 280 KLAIRATLKRSEFTFVFLGSALEKNGVPLDVALEVPNPESEVQVATILNKKDSKEKT 339
Db 242 KKGITKGCIMNSVPLVPLCGSSFKNGKVGQTLDDAVIDYLPAPTEVVDI---KGIDPKTEE 297
Qy 340 KILMNSSRHNSHPFVGLAPLEVGARF-GQLTVYRSYQGBLKKGDTIYNTTRKKVRLQRL 398
Db 298 EVFVKS--DDEFGALAFKIMTDPFVGQTFVVRVYRGKLESGSYVNSTKKKKVGRLL 355
Qy 399 AMHADMMASTEVEVYAGDICALFGI-DCASGDTFTDKANSGLSMSIHVPDPVISIMK 457
Db 356 LKMSHNRRE-DIKEVYAGSICAFVGLKDTLTGDTLDCDEKNA-VLERNMFPPEVHIHAYE 413
Qy 458 PSNKNDLEKFSKIGRFTREDPTFKVYPTDENKEVVISGMEHLIYAQRLEREYGCPC 517
Db 414 PRTKADQEKMGVALGKLAEDEBDFRVMTOEBTGQTLIGMELHLIYDRLKREKRYEA 473
Qy 518 ITGKRVAPRETITAPVPDFTHKKQSGAGQYGVYIGVLEPLDPEDYTKLEFSDETFGS 577
Db 474 EIGQPVAFRETIRRSVSKENHKYAKQSGRGQYGHVFIKLEPKEPG--SGYEFVNEISGG 531
Qy 578 NIPKQVPAVEKGFLLDAEKGKPLSGHKLGLRFLVODGAHMYDSNEISFIRAGSALKQ 637
Db 532 VLPKEVTPAVDGIQEAQNGVLAGYPVVDKFTLVDGSHYDVSSEMAFKIAGSVAFRKE 591
Qy 638 ALANATLILBEPIMAVEVAPNEFOGOVYAGINRRHGVITGQGVEDYFTLYADVPPLNDM 697
Db 592 ASRAANPVILBEPMMKVEVEVPEYMGDVIGDLNRRGQINSMDRLGLKIYVAFPLVEM 651
Qy 698 FGYSTELRSCTEGKGEYMEYSRYQPCLPSTOEDVINK 735
Db 652 FGYSTDLRSATQGRGTYSMEFPHYGEVPSNIAKEIYEX 689

RESULT 11
translation elongation factor EF-G - Helicobacter pylori (strain 26695)
C1Species: Helicobacter pylori
C1Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 02-Feb-2001
C1Accession: C64669
R1Tomu, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.; Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khakhria, H.G.; Glodek, A.; McKenna, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L. Nature 368, 539-547, 1997
A1Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C. A1Title: The complete genome sequence of the gastric pathogen Helicobacter pylori. A1Reference number: A64520; MUID:9739467; PMID:9252185
A1Accession: C64669
A1Status: preliminary; nucleic acid sequence not shown; translation not shown
A1Molecule type: DNA
A1Residues: 1-692 <TOM>
A1Cross-references: GB:AB000625; GB:AB000511; NID:g2314349; PIN:AA08239.1; PID:g231435
C1Superfamily: translation elongation factor G; translation elongation factor Tu homolog
C1Keywords: GTP binding; nucleotide binding; P-loop
F11-138/Domain: translation elongation factor Tu homolog <ETU>
F11-24/Region: nucleotide-binding motif A (P-loop)
F135-138/Region: GTP-binding NKXD motif

Query Match 38.1%; Score 1487; DB 2; Length 692;
Best Local Similarity 45.7%; Pred. No. 4,2e-89;
Matches 319; Conservative 129; Mismatches 232; Indels 18; Gaps 11;
Qy 42 PNEKIRNGISAHIDSGKTLTERVLYYTGRIAKNHEVKGXGVGAVNDMSLELROGIT 101
Db 6 PLNRIRNGIAAHIDAGKTTTERILFYGVSHKIGEV--HGG-AATYMDMDEKERTGT 62
Qy 102 IOSAATFTMKQDVNINIITDPGHVDFTEVERALRVLDGAVLVCAGVGOCOTMTVNRQ 161
Db 63 ITSAATFTMKQDVNINIITDPGHVDFTEVERALRVLDGAVLVCAGVGOCOTMTVNRQ 122
Qy 162 MKRYNVPFLFTINKLDRMGSNPARALQQRSKLNHNTAFVQIPMGLENGKGIVDLIEER 221

Db 123 ANKYGVPRIVFNKDKRIGANFNVENQIKRLKANPVPINPIGAEDTFIIVLQWAK 182
Qy 222 AIFYGDSQIARYG--EIPAEIAAADHRELIECVANSBOLEGEFLEKISISL 279
Db 183 ALVWNNK-TWGAAYDVEEIPSDLEKAKYREKLVAVAHFOEALMERTLGGELDIEI 241
Qy 280 KLAIRATLKRSEFTFVFLGSALEKNGVPLDVALEVPNPESEVQVATILNKKDSKEKT 339
Db 242 KKGITKGCIMNSVPLVPLCGSSFKNGKVGQTLDDAVIDYLPAPTEVVDI---KGIDPKTEE 297
Qy 340 KILMNSSRHNSHPFVGLAPLEVGARF-GQLTVYRSYQGBLKKGDTIYNTTRKKVRLQRL 398
Db 298 EVFVKS--DDEFGALAFKIMTDPFVGQTFVVRVYRGKLESGSYVNSTKKKKVGRLL 355
Qy 399 AMHADMMASTEVEVYAGDICALFGI-DCASGDTFTDKANSGLSMSIHVPDPVISIMK 457
Db 356 LKMSHNRRE-DIKEVYAGSICAFVGLKDTLTGDTLDCDEKNA-VLERNMFPPEVHIHAYE 413
Qy 458 PSNKNDLEKFSKIGRFTREDPTFKVYPTDENKEVVISGMEHLIYAQRLEREYGCPC 517
Db 414 PRTKADQEKMGVALGKLAEDEBDFRVMTOEBTGQTLIGMELHLIYDRLKREKRYEA 473
Qy 518 ITGKRVAPRETITAPVPDFTHKKQSGAGQYGVYIGVLEPLDPEDYTKLEFSDETFGS 577
Db 474 EIGQPVAFRETIRRSVSKENHKYAKQSGRGQYGHVFIKLEPKEPG--SGYEFVNEISGG 531
Qy 578 NIPKQVPAVEKGFLLDAEKGKPLSGHKLGLRFLVODGAHMYDSNEISFIRAGSALKQ 637
Db 532 VLPKEVTPAVDGIQEAQNGVLAGYPVVDKFTLVDGSHYDVSSEMAFKIAGSVAFRKE 591
Qy 638 ALANATLILBEPIMAVEVAPNEFOGOVYAGINRRHGVITGQGVEDYFTLYADVPPLNDM 697
Db 592 ASRAANPVILBEPMMKVEVEVPEYMGDVIGDLNRRGQINSMDRLGLKIYVAFPLVEM 651
Qy 698 FGYSTELRSCTEGKGEYMEYSRYQPCLPSTOEDVINK 735
Db 652 FGYSTDLRSATQGRGTYSMEFPHYGEVPSNIAKEIYEX 689

RESULT 12
translation elongation factor EF-G [imported] - Nostoc sp. (strain PCC 7120)
C1Species: Nostoc sp.
A1Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C1Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C1Accession: AC2348
R1Kanezo, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasaemoto, S.; Watanabe, A.; Iriyuchi, R.; Kanezo, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A1Title: Complete genomic sequence of the filamentous nitrogen-fixing Cyanobacterium Anabaena. A1Reference number: AB1807; MUID:21595285; PMID:11759840
A1Accession: AC2348
A1Status: preliminary
A1Molecule type: DNA
A1Residues: 1-692 <KUR>
A1Cross-references: GB:BA000019; PIDN:BA076037.1; PID:g17133474; GSPDB:GN00179
A1Experimental source: strain PCC 7120
C1Genetics:

Query Match 37.8%; Score 1472; DB 2; Length 692;
Best Local Similarity 44.9%; Pred. No. 4e-88;
Matches 319; Conservative 124; Mismatches 223; Indels 44; Gaps 11;
Qy 42 PNEKIRNGISAHIDSGKTLTERVLYYTGRIAKNHEVKGXGVGAVNDMSLELROGIT 101
Db 6 PLEKIRNGIAAHIDAGKTTTERILFYSGIHIGEV--HEGR-AVIDMDQERERIT 62
Qy 102 IOSAATFTMKQDVNINIITDPGHVDFTEVERALRVLDGAVLVCAGVGOCOTMTVNRQ 161
Db 63 ITAAISTSMQDVNINIITDPGHVDFTEVERALRVLDGAVLVCAGVGOCOTMTVNRQ 122

QY 162 MKRYNVPFLTFINKLDRMGSPARALQOMRSKLNHNHTAFWQIPMGLEGNFKGIVDLIEER 221
DB 123 ADKYKVPRIAFIKMORTGTANFYRHEQMDRLRANAIQIPISGENDFKGIIVDLVRKR 182
QY 222 AIYFDGDFSOIVRYGEIPAEIRAAATDHRQELIECVANSDEQL-----GEMFLEEKIPS 275
DB 183 AYYNNNDGDTDEETIIPADLQOVVEEYKIVAEVAETDDDLMSKYFPGSEPLTEERIS 242
QY 276 ISDLKAIIRATLKRSTPFLGSAKNGVOPLDVALEYLPNPSV-----QNA 327
DB 243 -----ALKRGITAGTIVPILGSAFKNKGVQMLDAVVDYLPAPTEVPIQCTLNGDA 296
QY 328 IANKDSDSEKRTKILNNSRHNSHPVGLAFPLEVGRFGOLTYVRSYOGELKKGDTTYNT 387
DB 297 IERRADNE-----PLAALAFKIMADPYGRGLTFVRVYSGVLKKGGSYVLA 341
QY 388 RTRKRVRLQRLAHMADMEASTEYVAGDICALFGI-DCASGDTFDKANSGLSMESIH 446
DB 342 TRKKKERISRLVLMKADDRQ-DVEELRAGDGLGALGLKDTLTGDTITDE-GAPVILESLF 399
QY 447 VDPDVLSIMKPSKNDLEKESKIGRTREDPTFKYVPTENKENVISGMGELHLEYA 506
DB 400 IPEPVLSVAVEPTKQDMDLKSLQSLSEDEPTFRVNDPETHQTVIAGMGLHLEILV 459
QY 507 QRLEREYGCPTCKPKVAFRETITAPV-FDFTHKQSGAGAGYKGVIGLEPLDEPY 565
DB 460 DRLRFRFYKANVANGAQAQVARETIRKPVNVBEKFIROSGGKQYGHVIVNLEPGRG-- 517
QY 566 TKLESEDEFSGNISIPKQFPAVKEGFLDACGKPLSGHKLGLRFLVLDGAHHMVDNSNI 625
DB 518 TGEFESKIVGVVPEXIGPAEQMKESCEGILGYPILIDYKATLVHGSYTDVDSSEM 577
QY 626 SFPRAGEGALKQALANATLCLEPIMAVEVAPNEGOGVAGINRRHGVITGQDVEY 685
DB 578 AFIASMAKKEVAVLKASPVLEPMKVEVEPEYDITGVIGDLISRRQGISQSTEGQL 637
QY 686 FTLYADVPINDMFGYSTELASCTEGKGEYTMESRYOCLPSTOEIVINK 735
DB 638 AKVASVPLATMGVATDIRSKTGGRGIFTMESHVEVPRSAVEITIAK 687

RESULT 13

AD0025
elongation factor G [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 17-May-2002
C:Accession: AD0025
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tlball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Taragde, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Dougan, G.;
11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AD0025
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-702 <KUR>
A:Cross-references: GB:AL50842; PIDN:CA089063.1; PID:g15978303; GSPDB:GN00175
C:Genetics: fusa
A:Gene: fusa
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog

Query Match 37.5%; Score 1462; DB 2; Length 702;
Best Local Similarity 44.4%; Pred. No. 1.8e-87;
Matches 317; Conservative 132; Mismatches 235; Indels 30; Gaps 15;
QY 42 PNEKIRNIGISAHIDSGKTLTERVLYYTGRIAKMHEVKGQGVAGVADSMELERQGIT 101
DB 6 PIERYRNIGISAHIDAGKTTTERILFYGVNHKIGEY--HDG-AATMDMEQEQERGIT 62
QY 102 IQSAATFTWKKV-----NINIITPGHVDFTIEVERALVLDGAVLVCAVGVOQ 154

DB 63 ITSAAITCFSWGMKQFEPHHVNIIDTPEHVDFTIEVERSMKRVLDGAVVYCAVGVOPO 122
QY 155 TWTVAROKMRYNVPFLTFINKLDRMGSPARALQOMRSKLNHNHTAFWQIPMGLEGNFKGI 214
DB 123 SETWRQANKYVPIAVVKNKDRMGANFLRVGGLKRLGANPVLQALGAEBKFTGI 182
QY 215 VDLIERAI-YFDGDFSOIVRYGEIPAEIRAAATDHRQELIECVANSDEQLGEMFLEEKI 273
DB 183 IDLVGMKALINWEADQVTFEYEEIPADMAELAEWQHNVLSAAEASDELMDKYLGESE 242
QY 274 ISDLKAIIRATLKRSTPFLGSAKNGVOPLDVALEYLPNPSVQNALINKD 333
DB 243 LTEEBIKALRQVLRKSEIILVTCGSAFKNKGVQMLDAVLEYLPAPTEV--SINGILD 300
QY 334 DSEKRTKILNNSRH--NSHPVGLAFPLEVGRF-GOLTYVRSYOGELKKGDTTYNTR 390
DB 301 DCKOTPAV-----RISDDKEPFSALAFKAIADPFQGNITFRVYSGIVNSGDTVANSYS 355
QY 391 KKVRLQRLAHMADMEASTEYVAGDICALFGI-DCASGDTFDKANSGLSMESIHVPD 449
DB 356 QERELGRIVQWYANKRE-EIKEVHAGDIAAIGLKDVTGTGLCD-PNNPIILERMEFPE 413
QY 450 PYISIAMKPSKNDLEKESKIGRTREDPTFKYVPTENKENVISGMGELHLEYAORL 509
DB 414 PVISVAVEPKTKADDEKMGMLGRLAKEDPSFRVWTDSESGTIIAGMGEHLIDLVRM 473
QY 510 BEYEGCPTCKPKVAFRETITAPV-FDFTHKQSGAGAGYKGVIGLEPLDEPYTKL 568
DB 474 RREFVEANVKGQVAYVETITRETYKDVGEKHAQSGRGQYGHVIVDMSPLPBG-VGY 532
QY 559 EESDETFSGNISIPKQFPAVKEGFLDACGKPLSGHKLGLRFLVLDGAHHMVDNSNISFI 628
DB 533 EFVNEIVGSGISIPKEPIPAVDKGIOEQKSGPLAGPVVDVAKRVLHGYSDHVDSSSLAFK 592
QY 629 RAGEBALKQALANATLCLEPIMAVEVAPNEGOGVAGINRRHGVITGQDVEDYFTL 688
DB 593 LAGSIAPKEGFRAPVLEPIMKVEVETPDYMGMDVGDINRRRGITLEGEDATATGCTV 652
QY 689 YADVLPINDMFGYSTELASCTEGKGEYTMESRYOCLPSTOEIVINKLEATGQ 742
DB 653 RYKVPLESEFAYIDLRSGTQGRASYSMEFLEYA-PS--NAKAVIEAR 702

RESULT 14

H81394
translation elongation factor EF-G Cj0493 [similarity] - Campylobacter jejuni (strain NC
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: H81394
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kestley, J.M.; Churcher, C.; Baaham, D.; Chillin
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell,
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp
A:Reference number: AB1250; MUID:20150912; PMID:10688204
A:Accession: H81394
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-691 <PAR>
A:Cross-references: GB:AL139075; GB:AL111168; NID:96967817; PIDN:CA075131.1; PID:9696796
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: fusa; Cj0493
C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog

Query Match 37.5%; Score 1460.5; DB 2; Length 691;
Best Local Similarity 44.3%; Pred. No. 2.2e-87;
Matches 309; Conservative 140; Mismatches 230; Indels 19; Gaps 11;
QY 42 PNEKIRNIGISAHIDSGKTLTERVLYYTGRIAKMHEVKGQGVAGVADSMELERQGIT 101
DB 6 PLKKVRNIGISAHIDAGKTTTSERILFTTGNSHKIGEY--HDG-AATMDMEQEQERGIT 62
QY 102 IQSAATFTWKKV-----NINIITPGHVDFTIEVERALVLDGAVLVCAVGVOQ 161

```

Db      63  ITSAATTCFMDHOHINTLIDTPGHVDFTEIEVERSMRLDQAVAFPCSVGVQOSSTVMQ 122
Qy      162  MKRYVVPFLTFINKLDRMGSNPAPALQOMRSKLNNHTAFQIPMGLEGFKIVLIEER 221
Db      123  ANKGVPRVVFVNKNDRIGANFYNNEDQIRNLKANPVPIQLPIGAEDNFKGVITDVTMK 182
Qy      222  AIYFDGDSQIVRY--GEIPAEIRAAATDHOELIECVANSDEQJGEMLEEKIPSIDL 279
Db      183  ALWEDD--TKPTDYVEKEIPAEIKKAEERYRTKMEIENAVSTSDLEMEKTLGGEELSLBI 241
Qy      280  KLAIRATLKASFTFVFLGSAALNKGVOPLDAVLELPLNPEVQNYAIIANKKDSKEKT 339
Db      242  KGIYAGCISLISIVMLCCTAFKNKGVOPLDAVVAIYLPAPDEVANI-----KGEYEDT 296
Qy      340  KILNNSRRHNSHPFVGLAFPLEVGR--GOLTYVRSYOGELKKGDITYNTRTKKYRLQRL 398
Db      297  EYVSAST--DGEFAGLAKIMTDPFVGOLTFVRVYRGCLSGSYAVANSTKCKEKIRGL 354
Qy      399  AMHADMMMEASTEVEYVAGDICALFGI--DCASGDTFTDKANSGLSMESIHPDPVISIAMK 457
Db      355  LKHGNNKHE--EIKVLYAGEIAGVVGKDTLTGDTLASBKIK--VILERMDFPDPVISVANE 412
Qy      468  PSNKNDLEKFSKIGRFTREDPTFKYYPFTENKEVYISGMGLHLEIYAQRLEREYGCPC 517
Db      413  PCTKADQEKMSIALNKLQEDPSFRVSTDEESGQTIISGMGLHLEIIVDRMLREPKYA 472
Qy      518  ITGKRVVAFRETITAPVPFDPFTHKKOSGAGQYGVKIVGLLEPLDEPDYTKLEFSDTEGS 577
Db      473  EYGOQVAVARETIRKTVBOEYRYAKOSGGRGQGHFRLPLEBEG--SGYFVNDIKGG 530
Qy      578  NIPKQFVPAVEKGPLDACEKGPLSGHKLSGRFLVLDGAFHMDSNELSFIRAGECALQ 637
Db      531  VIPKVIYPAVDGVOGEOALONGVLAGPVVEDVATVYDGSYHEVDSSEMAFKLAASMGFGE 590
Qy      638  ALAANATTLCLBPIYMAVEVVAENPEFGQVYIAGINRRHGVIYTGQDGVEDYFTLYADVPLNDM 697
Db      591  GARKGAVVLEBPMKVEVETPEDYMGDVIIGDLNKRQGVNSMDERGKNIITAFCEPLAEM 650
Qy      698  FGSTELRSCTEGKGEYTEMYSRYOCPCLSTOEDVINK 735
Db      651  FGYSTDLRSQOTGRATYSMEFPHYDEVPPKNAVDEIIRK 688

```

RESULT 15

S04390

translation elongation factor EF-G - *Spirulina platensis*C:Species: *Spirulina platensis*

C:Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 02-Feb-2001

C:Accession: S04390

R:Buttarelli, F.R.; Calogero, R.A.; Tiboni, O.; Gualerzi, C.O.; Pon, C.L.

Mol. Gen. Genet. 217, 97-104, 1989

A:Title: Characterization of the str operon genes from *Spirulina platensis* and their evolution

A:Reference number: S04388; PMID:89364697; PMID:2505055

A:Accession: S04390

A:Molecule type: DNA

A:Residues: 1-697 <BUT>

A:Cross-references: EMBL:X15646; NID:g47447; PIDN:CA33672.1; PID:g581727

C:Genetic8:

A:Gene: fus

A:Start codon: GTG

C:Superfamily: translation elongation factor G; translation elongation factor Tu homolog

C:Keywords: GTP binding; nucleotide binding; P-loop; protein biosynthesis

F.11-138/Domain: translation elongation factor Tu homology <ETU>

F.17-24/Region: nucleotide-binding motif A (P-loop)

F.135-138/Region: GTP-binding NKD motif

F.313-315/Region: GTP-binding SAK/L motif

F.23,24,62,135,138,313/Binding site: Mg-GTP (Lys, Thr, Asn, Lys, Asp, Ser) #acc

Query Match 37.3%; Score 1454; DB 2; Length 697;

Best Local Similarity 44.0%; Pred. No. 6e-877; Matches 309; Conservative 138; Mismatches 234; Indels 22; Gaps 11;

```

Qy      41  IPNEKIRNIGISAHIDSGKTLTERVLYYGRIAAMHEVKKQDGVAVMSDELERORGI 100
Db      5  IPLERVNIGIAAHIDAKTTTTERILFEYSGVHMGEV--HEGT-AVTDMWAQERRERGI 61
Qy      101  TIQSAATTTMKDVNINIIDTPGHVDFTEIEVERSMRLDQAVAFPCSVGVQOSSTVMQ 160
Db      62  TITAAISTSWLDHINIIDTPGHVDFTEIEVERSMRLDQAVAFPCSVGVQOSSTVMQ 121
Qy      161  OKRKYVVPFLTFINKLDRMGSNPAPALQOMRSKLNNHTAFQIPMGLEGFKIVLIEER 220
Db      122  QAEYQVRIAPFINKMDRTGADFFKVVQIHDRLRANAAPQVPPGERSDTHGLVDVAM 181
Qy      221  AIYFDGDSQIVRYG--EIPAEIRAAATDHOELIECVANSDEQJGEMLEEKIPSIDL 275
Db      182  KTYLYTNDLGTDIYQSDIEPREVOQLVAEYREKLEIENAVSTSDLEMEKTLGGEELSLBI 241
Qy      276  ISDLKLAIRATLKASFTFVFLGSAALNKGVOPLDAVLELPLNPEVQNYAIIANKKDS 335
Db      242  EEEIRHSLROGTIKGLIVPVICGSEFKNRGVORLDAVAVDLPAPTEVP--PIKGLVLDG 299
Qy      336  KEKTKILNNSRRHNSHPFVGLAFPLEVGRFGOLTYVRSYOGELKKGDITYNTRTKKYRL 395
Db      300  EBGVRYADDDA-----PLSALAFAKMAPDYGRLTFVRVYSGVLQKSYIYNATKAKKARI 354
Qy      396  QRLAMHADMMMEASTEVEYVAGDICALFGI--DCASGDTFTDKANSGLSMESIHPDPVISI 454
Db      355  SRLIYLKSD--EIRIEBELRAGDLGALGLKDTLTGDTICDEANS--IILESIYIPVISV 412
Qy      455  AMKPSNKNDLEKFSKIGRFTREDPTFKYYPFTENKEVYISGMGLHLEIYAQRLEREYGC 514
Db      413  AVEPKTKDMBELSKALQSLSEEDPTFRVSIIDSETNQTVIAGMGLHLEIIVDRMLREPK 472
Qy      515  CPCIRGKRVVAFRETITAPVPFDPFTHKKOSGAGQYGVKIVGLLEPLDEPDYTKLEFSDTE 574
Db      473  VEANIGAQVAVARETIRKTVBOEYRYAKOSGGRGQGHFRLPLEBEG--SGYFVNDIKGG 530
Qy      575  FGSNIPKQFVPAVEKGPLDACEKGPLSGHKLSGRFLVLDGAFHMDSNELSFIRAGECALQ 634
Db      531  VGSVPEKVIYPAVDGVOGEOALONGVLAGPVVEDVATVYDGSYHEVDSSEMAFKLAASMG 590
Qy      635  LKQALANATTLCLBPIYMAVEVVAENPEFGQVYIAGINRRHGVIYTGQDGVEDYFTLYADV 692
Db      591  IKNGVTKASPVYLEBPMKVEVETPEDYMGDVIIGDLNKRQGVNSMDERGKNIITAFCEPLA 650
Qy      693  PLNDMFGSTELRSCTEGKGEYTEMYSRYOCPCLSTOEDVINK 735
Db      651  PLATWFGYATIDRSKTQGRGVFSMEFPHYDEVPPKNAVDEIIRK 688

```

Search completed: July 14, 2003, 18:23:19

Job time : 19.8366 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:15:27 ; Search time 48.2242 Seconds
(without alignments)
4965.877 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730
Sequence: 1 FCLQSTRVWLRENGCHFPST.....MIVKKRYSTRSASSQSSSR 2057

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10730	100.0	2057	US-09-815-379-10	Sequence 10, Appl
2	10628	99.0	2058	US-09-815-379-17	Sequence 17, Appl
3	10247	95.5	2048	US-09-815-379-12	Sequence 12, Appl
4	5385	50.2	1031	US-09-764-868-647	Sequence 647, Appl
5	1789	16.7	348	US-09-764-868-1073	Sequence 1073, Appl
6	1789	16.7	348	US-09-955-999-101	Sequence 101, Appl
7	1492	13.9	2548	US-09-851-682A-1	Sequence 1, Appl
8	1461	13.6	1855	US-10-177-293-315	Sequence 315, Appl
9	1391	13.0	1285	US-10-205-823-273	Sequence 273, Appl
10	1391	13.0	1285	US-10-177-293-317	Sequence 317, Appl
11	1350	12.7	1016	US-10-044-303-2	Sequence 2, Appl
12	1333	12.4	1938	US-10-171-311-164	Sequence 164, Appl
13	1333	12.4	1972	US-10-171-311-162	Sequence 162, Appl
14	1329	12.4	765	US-10-044-303-1	Sequence 1, Appl
15	1328.5	12.4	1945	US-09-927-597-2	Sequence 2, Appl
16	1328.5	12.4	1979	US-09-927-597-4	Sequence 4, Appl
17	1327.5	12.4	2405	US-10-128-714-8280	Sequence 8280, Appl
18	1282	11.9	2099	US-10-128-714-3290	Sequence 3290, Appl
19	1206	11.2	845	US-09-927-597-14	Sequence 14, Appl

20	1201.5	11.2	849	9	US-09-927-597-10	Sequence 10, Appl
21	1201.5	11.2	852	9	US-09-927-597-12	Sequence 12, Appl
22	1114	10.4	772	9	US-09-927-597-8	Sequence 8, Appl
23	1012	9.4	699	9	US-09-927-597-6	Sequence 6, Appl
24	779	7.3	151	10	US-09-864-761-46352	Sequence 46352, A
25	779	7.3	151	10	US-09-864-761-48367	Sequence 48367, A
26	721	6.7	569	9	US-09-925-300-1583	Sequence 1583, Ap
27	713	6.6	569	9	US-10-106-698-5864	Sequence 5864, Ap
28	632	5.9	2568	10	US-09-866-108-3	Sequence 3, Appl
29	628	5.9	136	9	US-10-106-698-6839	Sequence 6839, Appl
30	549	5.1	570	10	US-09-764-853-551	Sequence 551, Appl
31	520	4.8	104	10	US-09-864-761-46427	Sequence 46427, A
32	487	4.5	94	10	US-09-864-761-48393	Sequence 48393, A
33	477	4.4	113	9	US-10-106-698-6068	Sequence 6068, Ap
34	388	3.6	80	10	US-09-864-761-46319	Sequence 46319, A
35	343	3.2	66	10	US-09-864-761-42111	Sequence 42111, A
36	313.5	2.9	245	10	US-09-925-302-686	Sequence 686, Appl
37	292.5	2.7	294	9	US-10-106-698-5501	Sequence 5501, Appl
38	288.5	2.7	2125	9	US-09-919-172-29	Sequence 29, Appl
39	264	2.5	53	9	US-09-764-891-2868	Sequence 2868, Appl
40	260	2.4	1239	9	US-09-291-417-13	Sequence 13, Appl
41	244.5	2.3	1233	9	US-09-291-417-89	Sequence 89, Appl
42	229.5	2.1	1175	10	US-09-771-161A-224	Sequence 224, Appl
43	229.5	2.1	1175	10	US-09-771-161A-225	Sequence 225, Appl
44	229.5	2.1	1175	10	US-09-771-161A-226	Sequence 226, Appl
45	228	2.1	240	9	US-10-101-487-75	Sequence 75, Appl

ALIGNMENTS

RESULT 1									
US-09-815-379-10									
; Sequence 10, Application US/09815379									
; Publication No. US20030073613A1									
; GENERAL INFORMATION:									
; APPLICANT: RASTELLI, LUCA									
; TITLE OF INVENTION: GERRITSEN, MARY									
; TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS									
; FILE REFERENCE: 10716/35									
; CURRENT APPLICATION NUMBER: US/09/815.379									
; CURRENT FILING DATE: 2001-03-22									
; PRIOR APPLICATION NUMBER: 60/191,134									
; PRIOR FILING DATE: 2000-03-22									
; NUMBER OF SEQ ID NOS: 17									
; SOFTWARE: Patentin Ver. 2.1									
; SEQ ID NO 10									
; LENGTH: 2057									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-09-815-379-10									
Query Match									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 2057; Conservative 0; Mismatch 0; Indels 0; Gaps 0;									
QY	1	FCLQSTRVWLRENGCHFPSTVNSCAEGIVPRRTDYGQFTYQSTTTHOKTANMPTRNE	60						
DB	1	FCLQSTRVWLRENGCHFPSTVNSCAEGIVPRRTDYGQFTYQSTTTHOKTANMPTRNE	60						
QY	61	GVDDNASITELHGSIMTNLFORYRGNQIMWTYGISILASVNPYOPJAGLYPATMEQYR	120						
DB	61	GVDDNASITELHGSIMTNLFORYRGNQIMWTYGISILASVNPYOPJAGLYPATMEQYR	120						
QY	121	RHLGSLPHIIPANECYRCILMKRDNOCTILKGSAGKTESTKLLIKFLSVISQOSLE	180						
DB	121	RHLGSLPHIIPANECYRCILMKRDNOCTILKGSAGKTESTKLLIKFLSVISQOSLE	180						
QY	181	LSLKRTSCVEKALIESSPIMEAFGNATVTYNNSSRGKPYQVNI CQKGNIOGGRIVDC	240						
DB	181	LSLKRTSCVEKALIESSPIMEAFGNATVTYNNSSRGKPYQVNI CQKGNIOGGRIVDC	240						

QY 241 ILSQNRVVRQNGERNYHIFVALLAGLBEHEBEFYLSTPENYHYLNOGCVEDKTI5D 300
DB 241 ILSQNRVVRQNGERNYHIFVALLAGLBEHEBEFYLSTPENYHYLNOGCVEDKTI5D 300
QY 301 QSFREVITAMDWQSKKEVREVSRLLAGILHGNIEFTTAGAQSFFTAGRAELL 360
DB 301 QSFREVITAMDWQSKKEVREVSRLLAGILHGNIEFTTAGAQSFFTAGRAELL 360
QY 361 GLDPTQUTDALTORSMFLRGEELITPLANOQAVDSDSLAMALYACCFEVIKINSRIK 420
DB 361 GLDPTQUTDALTORSMFLRGEELITPLANOQAVDSDSLAMALYACCFEVIKINSRIK 420
QY 421 GNEDFKSIIGLIDIFGFENEVNHFEQFNINANEKLOEYFNKHI FSLQLEYSREGLYWE 480
DB 421 GNEDFKSIIGLIDIFGFENEVNHFEQFNINANEKLOEYFNKHI FSLQLEYSREGLYWE 480
QY 481 DIDMINGECLDIIEKKGLILALINEESHFPQNTDSTLEKLSOHANNHFFYKPRVAVN 540
DB 481 DIDMINGECLDIIEKKGLILALINEESHFPQNTDSTLEKLSOHANNHFFYKPRVAVN 540
QY 541 NFGVKHAYAGEVOYDVARGILEKONDTFRDPLNLRSRPFYIDLFEHYSRRNODTLK 600
DB 541 NFGVKHAYAGEVOYDVARGILEKONDTFRDPLNLRSRPFYIDLFEHYSRRNODTLK 600
QY 601 GSKHRRPTVSSQFQKVDLSHSLMATTSSNPFVRCIKPNQKMPDQFQAVLNOQLRYSG 660
DB 601 GSKHRRPTVSSQFQKVDLSHSLMATTSSNPFVRCIKPNQKMPDQFQAVLNOQLRYSG 660
QY 661 METETVIRKAGVAVRRPFQDFYKRYVLMNLTALPBDVNGKCTSLQLIDANSSEWQOLK 720
DB 661 METETVIRKAGVAVRRPFQDFYKRYVLMNLTALPBDVNGKCTSLQLIDANSSEWQOLK 720
QY 721 TKVFLRESLEQKLEKREBEVSHAANVIRAHVGFAPLAKOYKVLVCVYIQNYAPFL 780
DB 721 TKVFLRESLEQKLEKREBEVSHAANVIRAHVGFAPLAKOYKVLVCVYIQNYAPFL 780
QY 781 RRRFLHLKKAIVFOKOLRGOIARVYROLLAEKREBEKKQKREBEKKREBEERER 840
DB 781 RRRFLHLKKAIVFOKOLRGOIARVYROLLAEKREBEKKQKREBEKKREBEERER 840
QY 841 ERREALRAQOEETKQOELALOKSOKAEITRELEKQKQOYBEILIRLEKEIDQ 900
DB 841 ERREALRAQOEETKQOELALOKSOKAEITRELEKQKQOYBEILIRLEKEIDQ 900
QY 901 RMKEQOELSTEA5LOKLOERRDOEIRLEBEACRAOEFLESINDEICRYNIERSI 960
DB 901 RMKEQOELSTEA5LOKLOERRDOEIRLEBEACRAOEFLESINDEICRYNIERSI 960
QY 961 SGGSEFSSSILAESACEKPNFNSQPYPEEVEDEGEADDAKDSPNPSEHGSHDQRTS 1020
DB 961 SGGSEFSSSILAESACEKPNFNSQPYPEEVEDEGEADDAKDSPNPSEHGSHDQRTS 1020
QY 1021 GRTSDSSSEEDPYMNDVVP7SPADSTVTLAPSVQDSGSLNSSGSESTYCMPONAD 1080
DB 1021 GRTSDSSSEEDPYMNDVVP7SPADSTVTLAPSVQDSGSLNSSGSESTYCMPONAD 1080
QY 1081 LPSPDGDYDODDYEDGATTSQSSVTFSNSYGSQSPVRCVGTYNSSGAIRFSSEG 1140
DB 1081 LPSPDGDYDODDYEDGATTSQSSVTFSNSYGSQSPVRCVGTYNSSGAIRFSSEG 1140
QY 1141 QSSFEDEEDFDSRFDITDEL5YRDSVYSCVTL7PYFHSFLYMKGGLMNSWKRMCVLD 1200
DB 1141 QSSFEDEEDFDSRFDITDEL5YRDSVYSCVTL7PYFHSFLYMKGGLMNSWKRMCVLD 1200
QY 1201 EFTFLMRSQOALKQMLHKKGGSSSTLSRNMKKMFWLRQSKMLYFENDSEKLGTV 1260
DB 1201 EFTFLMRSQOALKQMLHKKGGSSSTLSRNMKKMFWLRQSKMLYFENDSEKLGTV 1260
QY 1261 EYRTAKEI1DNTTKENGID1IMADRTEFH1AESPEDASQWFSVL5OVHASTDOEIOEMD 1320
DB 1261 EYRTAKEI1DNTTKENGID1IMADRTEFH1AESPEDASQWFSVL5OVHASTDOEIOEMD 1320
QY 1321 EOANPONAVGTL5VGLID5VCASDSPDRPNSFVITANRVLHCNADTPEBMHTMTLQOR 1380

DB 1321 EOANPONAVGTL5VGLID5VCASDSPDRPNSFVITANRVLHCNADTPEBMHTMTLQOR 1380
QY 1381 SKGDRVRGQEFIVGMHLKEVKN5PKMSLKLKRWPL7TNSLDYKSSSKNMLKGT 1440
DB 1381 SKGDRVRGQEFIVGMHLKEVKN5PKMSLKLKRWPL7TNSLDYKSSSKNMLKGT 1440
QY 1441 LVNL5CSVPPDEK1PKETGYMNTVYGRKHCYBL7YTLALNEATRWSSV1QNTDTKAP 1500
DB 1441 LVNL5CSVPPDEK1PKETGYMNTVYGRKHCYBL7YTLALNEATRWSSV1QNTDTKAP 1500
QY 1501 IDTPTQOLI0DIKENCLNSDVVEQIYKKNPILR7YTHPLHSL7LPLPGDINLNLKXG 1560
DB 1501 IDTPTQOLI0DIKENCLNSDVVEQIYKKNPILR7YTHPLHSL7LPLPGDINLNLKXG 1560
QY 1561 YTTLODEAKI1FNS1QOLESMSDPI1PIQGLIQTGHDLRPLDELYCOLIKQTNKVPHG 1620
DB 1561 YTTLODEAKI1FNS1QOLESMSDPI1PIQGLIQTGHDLRPLDELYCOLIKQTNKVPHG 1620
QY 1621 SVGNLYSQIL7CL5CTFLPSRGI7LYLKFHLKRIEOP7GTEMEKVALFTYESLKTTC 1680
DB 1621 SVGNLYSQIL7CL5CTFLPSRGI7LYLKFHLKRIEOP7GTEMEKVALFTYESLKTTC 1680
QY 1681 REFVPSRDEI5AL1HROEMTSTVYCHGGGCKIT7TNSHTAGEVVEKLI7RGLAMEDSRNM 1740
DB 1681 REFVPSRDEI5AL1HROEMTSTVYCHGGGCKIT7TNSHTAGEVVEKLI7RGLAMEDSRNM 1740
QY 1741 PALFENGVHVDKAI5R7TVADVADLAKFEKLA7TSEVGD7PMKFY7RCL7CF7LTDNVPKDS 1800
DB 1741 PALFENGVHVDKAI5R7TVADVADLAKFEKLA7TSEVGD7PMKFY7RCL7CF7LTDNVPKDS 1800
QY 1801 VEFAPMEQAEHVA1HGHHPAPEBN1QVLAAL7LO7YLO7DYTL7HA1P7LEVEYSIORLK 1860
DB 1801 VEFAPMEQAEHVA1HGHHPAPEBN1QVLAAL7LO7YLO7DYTL7HA1P7LEVEYSIORLK 1860
QY 1861 ARI5OSTKT7TP7CERLEKRR7S7LEGT7LR5R7TGS7VVRQ7KEE5OM7DM7KEEVSAR 1920
DB 1861 ARI5OSTKT7TP7CERLEKRR7S7LEGT7LR5R7TGS7VVRQ7KEE5OM7DM7KEEVSAR 1920
QY 1921 ASIIDKMRFOGMO7EO7MA7K7MALIKEMPGV7ST7FDV7CE7G67POE7LM7GV5ADAVS 1980
DB 1921 ASIIDKMRFOGMO7EO7MA7K7MALIKEMPGV7ST7FDV7CE7G67POE7LM7GV5ADAVS 1980
QY 1981 VYKRGEGRP7LEV7QYEH1LSFGAPLANTYKIVDERELL7FETSEVVDVAK7MKAYISMTV 2040
DB 1981 VYKRGEGRP7LEV7QYEH1LSFGAPLANTYKIVDERELL7FETSEVVDVAK7MKAYISMTV 2040
QY 2041 KRRYST7TNSASQSSR 2057
DB 2041 KRRYST7TNSASQSSR 2057
RESULT 2
US-09-815-379-17
; Sequence 17, Application US/09815379
; Publication No. US20030073613a1
; GENERAL INFORMATION:
; APPLICANT: RASTELLI, LUCA
; APPLICANT: GERITSEN, MARY
; TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE: 10716/35
; CURRENT APPLICATION NUMBER: US/09/815,379
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patent In Ver. 2.1
; LENGTH: 2058
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-379-17

Query Match	99.0%; Score 10628; DB 9; Length 2058;	
Best Local Similarity	99.3%; Pred. No. 0;	
Matches 2043; Conservative	5; Mismatches 7; Indels 2; Gaps 2;	
Qy	1 FCLQGTWVLRNGGHFPESTVNSCAGIYVFRTDVGQVFTYKQSTTHQKVTAMHPTNEE	60
Db	4 FTBEGRTVLRNGGHFPESTVNSCAGIYVFRTDVGQVFTYKQSTTHQKVTAMHPTNEE	63
Qy	61 GVDMAASLTLEHGGSIWNLFORYKRNQIWTYIGSTLASVNPQIPAGIYEPATMEQYR	120
Db	64 GVDMAASLTLEHGGSIWNLFORYKRNQIWTYIGSTLASVNPQIPAGIYEPATMEQYR	123
Qy	121 RHLAGELPHIIPAIANEVCYRCLMKRHDNQCILIKGESGAGCTESTKILKFLSVISQOSLE	180
Db	124 RHLAGELPHIIPAIANEVCYRCLMKRHDNQCILIKGESGAGCTESTKILKFLSVISQOSLE	183
Qy	181 LSLKETSCEVERAIILESSPIMEAFGNKTVYNNNSRFGKTVQANI CQKGNTOGGRIUDC	240
Db	184 LSLKETSCEVERAIILESSPIMEAFGNKTVYNNNSRFGKTVQANI CQKGNTOGGRIUDC	243
Qy	241 ILSQNRVVRONGERNYHIFYALLAGLHEERESEFYLSIPENYHILNOSGCEVEDTISD	300
Db	244 LIL-ENKRVVRONGERNYHIFYALLAGLHEERESEFYLSIPENYHILNOSGCEVEDTISD	302
Qy	301 QESFEVITAMVMOFSKEVEREVSRLAGILHLGNIEFITAGAQVSEKTAIGSAAEL	360
Db	303 QESFEVITAMVMOFSKEVEREVSRLAGILHLGNIEFITAGAQVSEKTAIGSAAEL	362
Qy	361 GLDPQLDALTORSWFLGEBILTPLANVOAVDSRDSI LAMALYACCFEWTIKTINSRIK	420
Db	363 GLDPQLDALTORSWFLGEBILTPLANVOAVDSRDSI LAMALYACCFEWTIKTINSRIK	422
Qy	421 GNEDEKSGIILIFGEBENFVNHFEQFINVANEXKOEYFNKHIFSLQLETYSRSEGLWE	480
Db	423 GNEDEKSGIILIFGEBENFVNHFEQFINVANEXKOEYFNKHIFSLQLETYSRSEGLWE	482
Qy	481 DIDWIDNGBECLDIEKKGLILALINEESHFPQATDSTLEKLSQHANNH FYVKPRVAVN	540
Db	483 DIDWIDNGBECLDIEKKGLILALINEESHFPQATDSTLEKLSQHANNH FYVKPRVAVN	542
Qy	541 NGVGVHAYGEVOYDVYRGILEKRDPTFRDLNLNLRSEKDFIYDLFEHVSRRNODTLKC	600
Db	543 NGVGVHAYGEVOYDVYRGILEKRDPTFRDLNLNLRSEKDFIYDLFEHVSRRNODTLKC	602
Qy	601 GSKHRRPTVSSQPKYDLSHSLMATLSSSNPFVRCIKPMQKMPDQDAVVLNQLRYSG	660
Db	603 GSKHRRPTVSSQPK-DSLSHSLMATLSSSNPFVRCIKPMQKMPDQDAVVLNQLRYSG	661
Qy	661 MLETVRIRKAGYAVRRPQDFYKRYKVLNRNLALPEDVHGKCTSLQLYDASNSEWQLGK	720
Db	662 MLETVRIRKAGYAVRRPQDFYKRYKVLNRNLALPEDVHGKCTSLQLYDASNSEWQLGK	721
Qy	721 TKVPLRESIEQLEKREBEVSHAAVIRAHVGLFARIQYRKVLCVVI IQNTYAFLL	780
Db	722 TKVPLRESIEQLEKREBEVSHAAVIRAHVGLFARIQYRKVLCVVI IQNTYAFLL	781
Qy	781 RRRFLHLKKAALVFOKOLAGOLARVYRQLAEKREBEKKEOESEKKEEERERER	840
Db	782 RRRFLHLKKAALVFOKOLAGOLARVYRQLAEKREBEKKEOESEKKEEERERER	841
Qy	841 EERRBAELRAQOEETRKQOELALQKQKEAELTRLEKQENKQYVEILRLKEKETEDQ	900
Db	842 EERRBAELRAQOEETRKQOELALQKQKEAELTRLEKQENKQYVEILRLKEKETEDQ	901
Qy	901 RKKEQOELSLTASLOKTOERRDOELRLLEBEACRAAQFSLSLNDEIDECVRNIERSL	960
Db	902 RKKEQOELSLTASLOKTOERRDOELRLLEBEACRAAQFSLSLNDEIDECVRNIERSL	961
Qy	961 SGGSEFSSFLASACEKPNFNSOPYPEEVEDEGEADDAFKOSPNSEHGHSQORTS	1020
Db	962 SGGSEFSSFLASACEKPNFNSOPYPEEVEDEGEADDAFKOSPNSEHGHSQORTS	1021

Qy	1021 GIRISDDSSSEEDPYNDITVPTSPSADSTVLAPSVODSGSLHNSGCEBETCMFQONAGD	1080
Db	1022 GIRISDDSSSEEDPYNDITVPTSPSADSTVLAPSVODSGSLHNSGCEBETCMFQONAGD	1081
Qy	1081 LPSPDGVDYDQDDVEDDAITSGSSVTFSNSVSGSOWSPDYRCSVCTYNNSSGAYRPSSEBA	1140
Db	1082 LPSPDGVDYDQDDVEDDAITSGSSVTFSNSVSGSOWSPDYRCSVCTYNNSSGAYRPSSEBA	1141
Qy	1141 QSSFEDSEEDPSRPTDDELSTYRDSVYSCVTLTYFHSFLYMKGGLHNSWRRRCVLDK	1200
Db	1142 QSSFEDSEEDPSRPTDDELSTYRDSVYSCVTLTYFHSFLYMKGGLHNSWRRRCVLDK	1201
Qy	1201 EFTLWFRSQEALXOGWLHKKGGSSTLSRRWKKRWFLRQSKLMYFENDSEBKLGCTV	1260
Db	1202 EFTLWFRSQEALXOGWLHKKGGSSTLSRRWKKRWFLRQSKLMYFENDSEBKLGCTV	1261
Qy	1261 EYRTAKEIINDNTKENGIDITIMADTTFHLIASEPDDASQWFSVLSQVHASTQOEIOEMND	1320
Db	1262 EYRTAKEIINDNTKENGIDITIMADTTFHLIASEPDDASQWFSVLSQVHASTQOEIOEMND	1321
Qy	1321 EQANPQNAVGTLDVGLIDSVCSADSPDRPNSFVITANRVLHCNADTPEBMMHWTLLOR	1380
Db	1322 EQANPQNAVGTLDVGLIDSVCSADSPDRPNSFVITANRVLHCNADTPEBMMHWTLLOR	1381
Qy	1381 SKGDTRVSGOEPIYVGMWLHKEVONS PKWSLKLKRWFLTNSLDYKSSBKNAKLGT	1440
Db	1382 SKGDTRVSGOEPIYVGMWLHKEVONS PKWSLKLKRWFLTNSLDYKSSBKNAKLGT	1441
Qy	1441 LVLNLSLCSVPPDEKIFKETGYWNTYVGRKRCYFLYTKLNEARMSVIONVTDTPAR	1500
Db	1442 LVLNLSLCSVPPDEKIFKETGYWNTYVGRKRCYFLYTKLNEARMSVIONVTDTPAR	1501
Qy	1501 IDTPTQOOLIODIKENCLNSDVYEQIYKKNPILRYTHHP LHSPLPLPYGDIMNLNLKDKG	1560
Db	1502 IDTPTQOOLIODIKENCLNSDVYEQIYKKNPILRYTHHP LHSPLPLPYGDIMNLNLKDKG	1561
Qy	1561 YTTLODEAKIKIPNSIQOLESNMSPPIPIIOGILQTHDRLPLDELYCQILKQTNKVPBHG	1620
Db	1562 YTTLODEAKIKIPNSIQOLESNMSPPIPIIOGILQTHDRLPLDELYCQILKQTNKVPBHG	1621
Qy	1621 SVGNLYSNQILTCLSLCTPLPSRGIILKYLKFLHKLRIEOPPGTBEKRYALPTYESLKTTC	1680
Db	1622 SVGNLYSNQILTCLSLCTPLPSRGIILKYLKFLHKLRIEOPPGTBEKRYALPTYESLKTTC	1681
Qy	1681 REFVPSRDEIBALIHROEMTSTVYCHGGGSCKITINSHTJAGEVYEKILRGLAMEDSRM	1740
Db	1682 REFVPSRDEIBALIHROEMTSTVYCHGGGSCKITINSHTJAGEVYEKILRGLAMEDSRM	1741
Qy	1741 FALPEYNGHVDAKIESTRVAVADVLAKEFKLAATSEVGLPMKFYFFLYCFDITDNDVPKDS	1800
Db	1742 FALPEYNGHVDAKIESTRVAVADVLAKEFKLAATSEVGLPMKFYFFLYCFDITDNDVPKDS	1801
Qy	1801 VEFAPMFQOAHAVIHGHHPABEENLOVLALRLLOYLOGDYTLHAAIPLLEBYVLSLQRIK	1860
Db	1802 VEFAPMFQOAHAVIHGHHPABEENLOVLALRLLOYLOGDYTLHAAIPLLEBYVLSLQRIK	1861
Qy	1861 ARISOSTKTPFCERLEKBRPTSFLGTLRSPRTGSVVRQKYEBSOMDMWIKKEVSSAR	1920
Db	1862 ARISOSTKTPFCERLEKBRPTSFLGTLRSPRTGSVVRQKYEBSOMDMWIKKEVSSAR	1921
Qy	1921 ASIIDKWRKFOGMOBQOMAKTMAIKEMPGYSTLPDVECKEGGPOELMIGVSADAVS	1980
Db	1922 ASIIDKWRKFOGMOBQOMAKTMAIKEMPGYSTLPDVECKEGGPOELMIGVSADAVS	1981
Qy	1981 VYKRGEGRPLEYFOYEHILSFGAPLANTYKIVYDERELLFETSEYVDVAKLMKAYISMTV	2040
Db	1982 VYKRGEGRPLEYFOYEHILSFGAPLANTYKIVYDERELLFETSEYVDVAKLMKAYISMTV	2041
Qy	2041 KKRYSSTRSASSQSSSR 2057	
Db	2042 KKRYSSTRSASSQSSSR 2058	

QY 1801 VEFAPMEQAEBAVIGHHHPAPEBNLQVLAALRLQYLOGDYTLHAIPLEEVYSLORLK 1860
DB 1801 VEFAPMEQAEBAVIGHHHPAPEBNLQVLAALRLQYLOGDYTLHAIPLEEVYSLORLK 1860
QY 1861 ARISSTKTFPCERLEKRTSFLBGLTARSFRITGSVVRCKYBEEQMLDMWIKKEVSSAR 1920
DB 1861 ARISSTKTFPCERLEKRTSFLBGLTARSFRITGSVVRCKYBEEQMLDMWIKKEVSSAR 1920
QY 1921 ASIIRKMRKFGOMNOEQAMAKYMALIKEMPGYSTLFDVECKEG 1964
DB 1921 ASIIRKMRKFGOMNOEQAMAKYMALIKEMPGYSTLFDVECKEG 1964
RESULT 4
US-09-764-868-647
; Sequence 647, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 647
; LENGTH: 1031
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (636)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-647
Query Match 50.2%; Score 5385; DB 9; Length 1031;
Best Local Similarity 99.5%; Pred. No. 1,2e-303;
Matches 1021; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
QY 1032 DRYNDVTVPSPADSTVLLAPSVODSGSLHNSGSESTYCMQONAGDLPSDDDYD 1091
DB 6 DAWNDVTVPSPADSTVLLAPSVODSGSLHNSGSESTYCMQONAGDLPSDDDYD 65
QY 1092 QDDYEDGATSSGSSVTFNSYGSQSPDYRCVGTNNSGAYRFSSEGAQSFEDSEDF 1151
DB 66 QDDYEDGATSSGSSVTFNSYGSQSPDYRCVGTNNSGAYRFSSEGAQSFEDSEDF 125
QY 1152 DSRPDTDELSTYRDSVYSCVTLPYFHSFLYMKGGLMSWKRWCVLKDETFLMFRSKOE 1211
DB 126 DSRPDTDELSTYRDSVYSCVTLPYFHSFLYMKGGLMSWKRWCVLKDETFLMFRSKOE 185
QY 1212 ALKQCMHLKKGSSSTLSRRNWKQWFLVROSKLMTFENDSEBKLKGYEVRTAKEIIDN 1271
DB 186 ALKQCMHLKKGSSSTLSRRNWKQWFLVROSKLMTFENDSEBKLKGYEVRTAKEIIDN 245
QY 1272 TTKENGIDILMDRFTFLAESPEBASQFVLSQVHASTDDEIEMHDEQANPQAVGT 1331
DB 246 TTKENGIDILMDRFTFLAESPEBASQFVLSQVHASTDDEIEMHDEQANPQAVGT 305
QY 1332 LDVGLIDSVACADSPDRPNSFVITANRVLHCNADTPEEMHMTITLLQSKSDTREVQOE 1391
DB 306 LDVGLIDSVACADSPDRPNSFVITANRVLHCNADTPEEMHMTITLLQSKSDTREVQOE 365
QY 1392 FIVRGMLHKEVKNSPKMSLKLKQWFLVTHNSLDYKSSSEKQALKGLVINSLSQVVP 1451
DB 366 FIVRGMLHKEVKNSPKMSLKLKQWFLVTHNSLDYKSSSEKQALKGLVINSLSQVVP 425
QY 1452 PDEKIFKETGYNNVTYGRKHCYRLTYTKLNEATRRSVIOWNTDTPKATIDTPTQOLIOD 1511
DB 426 PDEKIFKETGYNNVTYGRKHCYRLTYTKLNEATRRSVIOWNTDTPKATIDTPTQOLIOD 485
QY 1512 IKENCLNSDVEBOIYKRNPIILRYTHPLHSPLLPLPYGDIINLNLKDKGYTTLQDEAIKI 1571

DB 486 IKENCLNSDVEBOIYKRNPIILRYTHPLHSPLLPLPYGDIINLNLKDKGYTTLQDEAIKI 545
QY 1572 FNSLOQESMSDPIPIIOGILQTHDLRLADELYCOLIKOTNNVPHFGSGNLYSMQIL 1631
DB 546 FNSLOQESMSDPIPIIOGILQTHDLRLADELYCOLIKOTNNVPHFGSGNLYSMQIL 605
QY 1632 TGLSCTFLPSRGLIKYLLKFLKRIREQPPGEMEKYALFTYESLJKTCKRFPVSRDIE 1691
DB 606 TGLSCTFLPSRGLIKYLLKFLKRIREQPPGEMEKYALFTYESLJKTCKRFPVSRDIE 665
QY 1692 ALIHRQEMSTVYCHGGSGCKITINSHTTAGEVVEKILRGLAMEDSENMPALPEYNGHD 1751
DB 666 ALIHRQEMSTVYCHGGSGCKITINSHTTAGEVVEKILRGLAMEDSENMPALPEYNGHD 725
QY 1752 KAISRTVAVDLAKFEELATSEVGDLPKPFYKLVYCFDLDVNPQKOSVAPAFMEQAH 1811
DB 726 KAISRTVAVDLAKFEELATSEVGDLPKPFYKLVYCFDLDVNPQKOSVAPAFMEQAH 785
QY 1812 EAVIGHHPABEENLQVLAALRLQYLOGDYTLHAIPLEEVYSLORLKARISOSTKFT 1871
DB 786 EAVIGHHPABEENLQVLAALRLQYLOGDYTLHAIPLEEVYSLORLKARISOSTKFT 845
QY 1872 PCRLEKRTSFLBGLTARSFRITGSVVRQKYBEEQMLDMWIKKEVSSARASIIDKMRKQ 1931
DB 846 PCRLEKRTSFLBGLTARSFRITGSVVRQKYBEEQMLDMWIKKEVSSARASIIDKMRKQ 905
QY 1932 GMDQEQAMAKYMALIKEMPGYSTLFDVECKEGFPOLMIGVADAVSVYKRGGRLE 1991
DB 906 GMDQEQAMAKYMALIKEMPGYSTLFDVECKEGFPOLMIGVADAVSVYKRGGRLE 965
QY 1992 VFQYEHILSPAPLANTYKIYVDERELFETSEVVDVAKLAKAYISHIWKRYSTTRRAS 2051
DB 966 VFQYEHILSPAPLANTYKIYVDERELFETSEVVDVAKLAKAYISHIWKRYSTTRRAS 1025
QY 2052 SQGSSR 2057
DB 1026 SQGSSR 1031
RESULT 5
US-09-764-868-1073
; Sequence 1073, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 1073
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1073
Query Match 16.7%; Score 1789; DB 9; Length 348;
Best Local Similarity 99.1%; Pred. No. 5.6e-96;
Matches 335; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1032 DRYNDVTVPSPADSTVLLAPSVODSGSLHNSGSESTYCMQONAGDLPSDDDYD 1091
DB 6 DAWNDVTVPSPADSTVLLAPSVODSGSLHNSGSESTYCMQONAGDLPSDDDYD 65
QY 1092 QDDYEDGATSSGSSVTFNSYGSQSPDYRCVGTNNSGAYRFSSEGAQSFEDSEDF 1151
DB 66 QDDYEDGATSSGSSVTFNSYGSQSPDYRCVGTNNSGAYRFSSEGAQSFEDSEDF 125
QY 1152 DSRPDTDELSTYRDSVYSCVTLPYFHSFLYMKGGLMSWKRWCVLKDETFLMFRSKOE 1211

Db 126 DSRFDTDELRYRDSVYSCVTLPRFHSFLYMKGLMNSWKRRWCYLKDETFLMFRSKOE 185
 QY 1212 ALKQGLHKKGGSSSTLSRRNMKKRWFVLROSKLMTFENDSEKLGTVETRAKTIIN 1271
 Db 186 ALKQGLHKKGGSSSTLSRRNMKKRWFVLROSKLMTFENDSEKLGTVETRAKTIIN 245
 QY 1272 TTKENGIDIIIMADRTFLIAESPEDASQMFVSLOVHASTDOEIOEMHDEQANPQNAVGT 1331
 Db 246 TTKENGIDIIIMADRTFLIAESPEDASQMFVSLOVHASTDOEIOEMHDEQANPQNAVGT 305
 QY 1332 LDVGLIDSVCASDSPDRPNSFVITITANRVLHCNADTPE 1369
 Db 306 LDVGLIDSVCASDSPDRPNSFVITITANRVLHCNADTPE 343

RESULT 6

US-09-955-999-101
 ; Sequence 101, Application US/09955999
 ; Publication No. US20030036505A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Barash et al.
 ; TITLE OF INVENTION: Signal Transduction Pathway Component Polynucleotides, Polypeptide
 ; TITLE OF INVENTION: Antibodies, and Methods Based Thereon
 ; FILE REFERENCE: FT066P1
 ; CURRENT FILING DATE: 2001-09-20
 ; PRIOR APPLICATION NUMBER: 60/234,997
 ; PRIOR FILING DATE: 2000-09-25
 ; NUMBER OF SEQ ID NOS: 139
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 101
 ; LENGTH: 348
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-955-999-101

Query Match 16.7%; Score 1789; DB 9; Length 348;
 Best Local Similarity 99.1%; Pred. No. 5.6e-96;
 Matches 335; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1032 DRYMNTTVPTSPASSTVILLASVDDSGSLHNSGSESTYCMPOAGDLPSPDGDYD 1091
 Db 6 DAWVNDTVPTSPASSTVILLASVDDSGSLHNSGSESTYCMPOAGDLPSPDGDYD 65
 QY 1092 QDDYEDGATISGSSVTFNSYSGQSPDYRCVGTNNSGAYRFSSEGAQSSPEDSEDP 1151
 Db 66 QDDYEDGATISGSSVTFNSYSGQSPDYRCVGTNNSGAYRFSSEGAQSSPEDSEDP 125
 QY 1152 DSRFDTDELRYRDSVYSCVTLPRFHSFLYMKGLMNSWKRRWCYLKDETFLMFRSKOE 1211
 Db 126 DSRFDTDELRYRDSVYSCVTLPRFHSFLYMKGLMNSWKRRWCYLKDETFLMFRSKOE 185
 QY 1212 ALKQGLHKKGGSSSTLSRRNMKKRWFVLROSKLMTFENDSEKLGTVETRAKTIIN 1271
 Db 186 ALKQGLHKKGGSSSTLSRRNMKKRWFVLROSKLMTFENDSEKLGTVETRAKTIIN 245
 QY 1272 TTKENGIDIIIMADRTFLIAESPEDASQMFVSLOVHASTDOEIOEMHDEQANPQNAVGT 1331
 Db 246 TTKENGIDIIIMADRTFLIAESPEDASQMFVSLOVHASTDOEIOEMHDEQANPQNAVGT 305
 QY 1332 LDVGLIDSVCASDSPDRPNSFVITITANRVLHCNADTPE 1369
 Db 306 LDVGLIDSVCASDSPDRPNSFVITITANRVLHCNADTPE 343

RESULT 7

US-09-851-682A-1
 ; Sequence 1, Application US/09851682A
 ; Patent No. US20020091248A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Adams, Arwen E.
 ; APPLICANT: Chiu, Choi Yung
 ; APPLICANT: Duhl, David

; APPLICANT: Gorman, Susan W.
 ; APPLICANT: Leng, Song
 ; APPLICANT: Sheffield, Val
 ; APPLICANT: Welch, Juliet
 ; TITLE OF INVENTION: MYOSIN IYA AND CYCLIC NUCLEOTIDE GATED
 ; TITLE OF INVENTION: CHANNEL-15 (CNGC-15) POLYNUCLEOTIDES, POLYPEPTIDES,
 ; FILE REFERENCE: 200130.442
 ; CURRENT FILING DATE: 2001-05-08
 ; PRIOR APPLICATION NUMBER: US/09/851,682A
 ; PRIOR FILING DATE: 1998-10-14
 ; NUMBER OF SEQ ID NOS: 3
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 1
 ; LENGTH: 2548
 ; TYPE: PRT
 ; ORGANISM: Homo sapien
 US-09-851-682A-1

Query Match 13.9%; Score 1492; DB 10; Length 2548;
 Best Local Similarity 24.3%; Pred. No. 1.2e-77;
 Matches 541; Conservative 349; Mismatches 674; Indels 660; Gaps 79;

QY 35 YGQVETTKQSTITHQKTA--MPTNBEGVDDMASLTGLHGSIMYNLFQRYKKNQIWT 91
 Db 118 YGSLQSWLRVTERRRMRMERGFLPQPOQKDFDLCSLPDLNKEKTLLENLRDRFKGKIY 177
 QY 92 YIGSLASVNPQPIAGLYEPATMEQYRRHGLPRHIFAIANECYCLMKRHNQCIL 151
 Db 178 YVGSILIVINPEK-FLPIYNPYVMYDNHOGKEPPIYVADVAYAMQKKNQCI 236
 QY 152 IKESGAGTSTKILKFLSVISQOSLELSKEKTCSEARILESSPIMEAFNAKTVY 211
 Db 237 ISGESGSGKTGNTNLHLTLRSQGF-----ASGEQIILGAGVLEHFGAKTAH 289
 QY 212 NNNSSRFQFVQNLICQKNIQCGRIVDICLSSQNRVVRQNGERNYHIFALLAGLEHE 271
 Db 290 NNNSSRFQFQVNVQETGVLA-YVEKYLLEKRLVYQEHNEBNYHVFYLLAGASD 348
 QY 272 EREBYLSTPENYHNLN-----SGVEDEK-----TISDOE--SPREVTAMDV 313
 Db 349 EBSAFLKQPEBYHNLNITKKPLRQSWDYCYSEPDCTFVGEGLDLDHDERLQLANEM 408
 QY 314 MOPSREBYREVSRLAGILHGN-----EFTAGGAOVSPYALGRSABELGIDPTOLT 369
 Db 409 VGFLEKTRQIFSLSALILHGNICYKKKTRDSDIDCNPEVLPVSEELLEVEKEMLPE 468
 QY 370 ALTORSMEFLRGEELTPLVQOAVDSRDSLAMALYACCFEYVYKKNIRKGNEDFK-- 426
 Db 469 ALVTBKTVTGBEKLIPYKLABAVTVRNSMAKSLYALFDMIVFRINHALNSKOLEHNT 528
 QY 427 ---SIGIIDIGFEFEFVNNHFEQPINYANKEQLQYFKNHIFSLQLEYSRGLWEDID 483
 Db 529 KTLSTGIVDIFQFEDYENNSFEQFICINANRLOQYFQOHFKLEQEBYRREGISWNI 588
 QY 484 WIDNECDLLEK-LGLALINESHFPOATDSTLLEKLSOAHNNFYVYKPAVANNF 542
 Db 588 YIDNCTCNILSKRTGLHLHIDESNFPQATNOTLLKFKQHEDNSYIEFPVAMEPAF 648
 QY 543 GVKHYAGEVOYDVRCILEKNRDTFRDDLNLNLRSPDFIYL----- 585
 Db 649 IIKHYAGVKYGVQDFREKNTDHRKPDIVALLRSKNFISGMIGIDPVAVFMALRAF 708
 QY 586 -----FHVSSRN-----NODTLKCG----- 601
 Db 709 FRAMVAFREAGKRNIRKTHGDDTAPCAILKMSDSFSLQHPVQRSLLEIQRCKERY 768
 QY 602 ----- 601
 Db 769 ITRKNPRTPLSDQSMALNEKNQHTFDIAMNGRTGIRQSLSSGTSLLDKDGFANST 828
 QY 602 -----SKHRRP 607

Db	829	SSKLLERAHGILTRNKNFKSPALPKHLLVNSLKHLLRLTLQDRITKSLHLHKKKKPP	888
Qy	608	TVSSQFKDLSHLSMATSSSNPFVRCIKPMQMPQOFDAVVLNOLRISGHELTRI	667
Db	889	SISQFOQA-SLSKMETETGOAEPYFVKCIRSNAEKLPIRFSVDVLRQLRYATGMLETQOI	947
Qy	668	RKAGAAVRPPRODFKRYKLYMRNLALBEDYRGKTSLLGLYDANSNSMOGKTVFVRE	727
Db	948	ROSGVSKSYSDQDFVSHFVHLPPNITPS--KPNIQDFPKTNLPDPNYQGKTMVFLKE	1005
Qy	728	SLEQKLEKREBEV-----SHAMVIRAHVLTGLFARQYVK-	763
Db	1006	QERHLODLHNOELVARIILLOQMRVLLSCQHFLHLQFASIIQRPFRNYLNOQVYDA	1065
Qy	764	-----VLVCVVIIOKNYPAFLRLRRFLHKKAIVFCQLARQIARR-----	805
Db	1066	AVOKDAFYMAAALLQASWRAHLERQYULELAAAIYICQKWDYVRRRHMAIICIQAR	1125
Qy	806	--VYROLLAEKREQEKKKOBEKKREBEERERERERFLELAAQOE-----ETRK	857
Db	1126	WKAYRE--SKRYOQRKKIILLOSTCGFRARQFKALFKQRLRETRPELVNIXQY	1182
Qy	858	QOELEALQSQ-KEAELTRELEKOKENQVEIEILYELKEDLOQMK-----QOELSTGE	912
Db	1183	SLEIOGSPSEWEDOSPDKRIKATECKSVIESNFISE--SSVDCKKSPKQDRAASQ	1241
Qy	913	ASLOKLOE-----RRDOELRLLEEACRAOEFLESLNFEIDECEVRNIE-----RSLSGG	963
Db	1242	SGVD-LQGDVLYREPRSLLEDHOKQVBARAK--ESRMFELEQAIPLSELLKVRSLGI	1298
Qy	964	S-----EFSSELABACEE--KPNFNSQ-----PYREE-----VDEGREADD	1001
Db	1299	SPSEDRRSTELVPEGLQSPGTPSPSSQSLBELSYECSKLSIEVISBEG-----D	1353
Qy	1002	AFKQSPNSEGHNSPRTSGIRTPDQSE-----EDPYNDVTVPPSPADSVTLAPSVQ	1054
Db	1354	LQFSPKLISSSPKFDSDRNALASNETISAEHLKQTKMKNVCSSES---ITCKPOLK	1406
Qy	1058	DGSLHNSGSESTYCMPONAGDLPSPDGDYDODDYEIGALITSGSVTFSNSYGSQMS	1117
Db	1410	DS---PISNLSPTFPYIIOQOBLKTNQSLDTSIQNKLLHEDTIGALTLIDINETR-	1466
Qy	1118	PDYRCS-----VGTYNSSGAYRFSSEGAQSSFEDEEDFUSRPTDDEL-----	1161
Db	1465	-RYHCSGKQIYVPSLINTESSNPVLKLEKLTNTEKEBERQKIQOQNEKEMEQIROQTDL	1523
Qy	1162	SYRDSYVSCVTLPEYFISPLMKGGLMSMKRKCVCYLKDETFLMFRKQE-	1211
Db	1524	EKERKAFYTBKPRJGCELVAP-----SSYQSKQRYVRPSSLTSLNTSNKGE	1570
Qy	1212	-----ALKQOMLHKGGGSSSTLSRRMKKRMFVLRQSLIMYFENDSEBKLVGV---	1266
Db	1571	LNVLGSLSLKAAALLQAKSSSAHLPPKD-----RPVTVFERKKSPOCSSTVYKLS	1621
Qy	1261	-----EVRTAKEIIDN-----TTKEN-GIDI	1286
Db	1622	KTRMGTOVLNAACKLSNNRISREHFRPTQSYSHNSDDLREGNARPIFFPPKQWMSJPL	1681
Qy	1281	IMADRTFLILIESP---EDASQW--FSVLSQVHASTQOUIOEMHDEQANPONAVGTLDV	1334
Db	1682	VSKEA--LNSKNPDLHKEDEBAMRPVVLACGQRETCQRFSSV--DEQAKLHK--TWSQ	1734
Qy	1335	GLIDBVC-----ASDQPRPNPSPVILITARVYLHCNADTBEIMHNYTLLOSQSGTRYE-	1388
Db	1735	GEITTLAVROKASDSDIRPQ-----ALORFMAKKGQGEKTKTRVXPT	1777
Qy	1389	-----QOEFI-----VRGVLHKEYONSPKMSLTKKRW	1417
Db	1778	TQSEVSPPLFAGTDVILPAHQPDLELAANYHTPRLSPELQSCKEKEKEMKEPSP-----	1833
Qy	1418	FVLTHNSLDYKSSBKALKGLTVLNSICSVVPDEKILKETGYMNVTVYGRKHCYRL	1477

```

Db 1831 -----KAKRKSXKSNVALDSM-----HMW-----1851
Qy 1478 TKLINEKTRMSVQNTOTDKAPIDTPTQOULIODI-KENCINSDVVEQIYKRN-PILRYT 15355
Db 1852 ----NDSSVQ-----IIASVSJLKS-MDEFILKKVNDLDMEDSFKODLVAVVPFKAKKEFRQN 19034
Qy 1536 HHPLHSPULPLPYDIDINLNLKXGTYTTLQDEAKIKRNSLDQLESMD-PPIIQLIQT 1594
Db 1904 IFSPYSALAMDGD-----KSIRYKDYALPEOLIEKTRLEORDSLGESPFRV-----W 19533
Qy 1595 GHDIRPLRDELYCOLLIQOTN-----KVP-----HPESVGLYSMOULT- 1632
Db 1954 VNTPEKVLDE-YMNEFTSDCTATKVPKTERKKRKETDLVEBHNGHIFRATYSLPTY 2012
Qy 1633 CLSCTPLPSRGILKYLKFKHLKRIEOPFGTEMEKYALFTYESLKT-KC-----R 1661
Db 2013 CEYCS-----SLIWMIDRASVCLCYACHKKCCLTTPYAKCSKKYDPELSSR 20595
Qy 1682 EF-----VPSRDE-----IEALIHROBM-----TSTVYCHGSGSCKITINSHTAGBVE 1726
Db 2060 QFGVLSRLTSSEDTPLVVEKLIINYEMHGLYEGYIRKSGSTNKI-----K 2107
Qy 1727 KLIRGLAME-DSRMFALFEYNGHVDAKIESRTVAVDLAKFEKILAATSBEVDLPWKPY- 1784
Db 2108 ELRGGLTDAESVN---LDDYNIH-----VIASVFKOM-----LRDLPNPLMT 2147
Qy 1785 FKLY 1788
Db 2148 FELY 2151

```

```

, RESULT 8
, US-10-177-293-315
, Sequence 315, Application US/10177293
, Publication No. US20030124126A1
, GENERAL INFORMATION:
, APPLICANT: Lillie, James
, APPLICANT: Glatt, Karen
, APPLICANT: Zhao, Xumei
, APPLICANT: Gannavarpu, Manjula
, APPLICANT: Kamatekar, Shubhangi
, APPLICANT: Mertens, Maureen
, APPLICANT: Myer, Vic
, APPLICANT: Wang, Youzhen
, APPLICANT: Xu, Yongyao
, APPLICANT: Hoersch, Sebastian
, APPLICANT: Monahan, John
, APPLICANT: Meyers, Rachel E.
, APPLICANT: Bat Jr., Robert C.
, APPLICANT: Hortobagyi, Gabriel N.
, APPLICANT: Puzatzi, Lajos
, APPLICANT: Meric, Funda
, APPLICANT: Sahin, Aysegul
, APPLICANT: Mills, Gordon B.
, TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT
, TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
, FILE REFERENCE: MRI-038
, CURRENT APPLICATION NUMBER: US/10/177,293
, CURRENT FILING DATE: 2002-06-21
, PRIOR APPLICATION NUMBER: US 60/299,887
, PRIOR FILING DATE: 2001-06-21
, PRIOR APPLICATION NUMBER: US 60/301,572
, PRIOR FILING DATE: 2001-06-27
, PRIOR APPLICATION NUMBER: US 60/306,501
, PRIOR FILING DATE: 2001-07-18
, PRIOR APPLICATION NUMBER: US 60/325,002
, PRIOR FILING DATE: 2001-09-25
, PRIOR APPLICATION NUMBER: US 60/362,585
, PRIOR FILING DATE: 2002-03-05
, PRIOR APPLICATION NUMBER: US 60/xxx,xxx
, PRIOR FILING DATE: 2002-05-14
, NUMBER OF SEQ ID NOS: 506
, SOFTWARE: FastSeq for Windows Version 4.0

```



```

RESULT 9
US-10-205-823-273
; Sequence 273, Application US/10205823
; Publication No. US20030108963A1
GENERAL INFORMATION:
APPLICANT: Schlegel, Robert
APPLICANT: Monahan, John E.
APPLICANT: Endege, Wilson O.
APPLICANT: Gannavarapu, Manjula
APPLICANT: Gorbacheva, Bella
APPLICANT: Hoersch, Sebastian
APPLICANT: Kamackar, Shubhangi
APPLICANT: Monsey, Angela M.
APPLICANT: Glatz, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Anderson, Duetin
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
FILE REFERENCE: MRI-044
CURRENT APPLICATION NUMBER: US/10/205,823
CURRENT FILING DATE: 2002-07-25
PRIOR APPLICATION NUMBER: 60/307,982
PRIOR FILING DATE: 2001-07-25
PRIOR APPLICATION NUMBER: 60/314,356
PRIOR FILING DATE: 2001-08-22
PRIOR APPLICATION NUMBER: 60/325,020
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: 60/341,746
PRIOR FILING DATE: 2001-12-12
PRIOR APPLICATION NUMBER: 60/362,158
PRIOR FILING DATE: 2002-03-05
NUMBER OF SEQ ID NOS: 455
SOFTWARE: PaedSeq for Windows Version 4.0
SEQ ID NO 273
LENGTH: 1285
TYPE: PRF
ORGANISM: Homo sapiens
US-10-205-823-273

Query Match      13.0%; Score 1391; DB 9; Length 1285;
Best Local Similarity 31.1%; Pred. No. 3,4e-72;
Matches 368; Conservative 225; Mismatches 365; Indels 226; Gaps 33;

QY 10 LBNGHFPSTYNSCAEIVFRITDYGQVYTKOSTITTHQKTAHMFTEBEGDDMAST 69
DB 34 LKQKGTPLALIN-----QVPAEED-----SKQVEDNCSLM 66
QY 70 ELHGSIMYNLPQRYKRNQIMWTYIGSILASVNPYQPIAGLYEPATWEOYSRRHLGELPH 129
DB 67 YLNEATLNLINIKRYSKDRIYTVVANILAVNPYPIPIKIYSEAKISYQGSLSGRPH 126
QY 130 IPALANECCYRLKWRHNOCILKGSAGKTESTYLILKFLSVISQSLSLKERTSC 189
DB 127 VFALADKARFDMKVLKMSQSIIVSGSGAGKTEMTKFLRYLTESYGTGD----- 177
QY 190 VERALLESSPIMEAFGNACTVYNNSSRGKPVQANI:QKGNIGGRYIDCLISSQNV 249
DB 178 IDRIVEANPPLLEAFGNACTVNNSSRGKPVFIHFNRKSSVAGD-FVSHYLLKSRIC 236
QY 250 RONGERNYHIFAYLAGLEHEEREFYSTEPNYVYLN----- 288
DB 237 VQKERNYHIFAYRLCAGSBDIREKHLSSPDNPFYLRGCTRYRANKETDKOILQNRK 296
QY 289 -----QSGCEDTKTISDQSFREVTAMDVMOFSKEEVEBSLLAGILHGNIEF--- 339
DB 297 SPEYLAGSMKDPILDDHGDFFIMCTAMKKIGLDBEKLDFRVVAGVLAHLNDPEBAG 356
QY 340 ITAGAGVQVFKTA--LGRSAILGLDPTQLTALTLQFSMF-----LRGEIILPLNVQA 392
DB 357 STSNGCNLKNKSKQSLEYCAELIGLDQDDRLVSLTTRVNLTTAGGTGTIVIKVPLVEQA 416

```

```

QY 393 VDSRDSLAMALYACCFFENVIKINSRIGKNEDPKSIIGLIDIFGFENFEVNHFEQFNINYA 452
DB 417 NNARDALAKTVYSHLFDHVVNRVNCPPFETSYRIFGLVLDIAGFEYFEHNSPEQCINVC 476
QY 453 NEKOEYNNKXIFSLQCEYSREGVWEDIDIMDGECLDILIEKL-GLLALINESHP 511
DB 477 NEKQOFFNERILKEBOELYOKEGIVNHYVNDQCIDILIBALTVOILIDLBENLIP 536
QY 512 QATDSTLEKXHSOAHNHFYV---KPRVNV-----NPGVGYAGEVOYDVRGILE 560
DB 537 QPSDQHTSAVHQK-KHFRLLTTRKSKLAVHRIHRIHDEGFIIRHAGANCYETTORVE 595
QY 551 KNKRTFDRDLNLLESREFDIYDLFEHVSRNNDTLKCSKRRPVTVSQFVDSLHS 620
DB 596 KNNDALHMSLESBLESKDFIRELFES-STNNKNDTQKXKGLSFISVGNKFKT-QNL 653
QY 621 LMATSSNPFVRCIKRNQMOPQOAVVLNOLRISGMLETVIRIKAGYAVARRPQD 680
DB 654 LLDKLRSTGASFIRCIKRNKMTSHHFEQAQILSQCSGNVSVYDLMOGGYPSRASFHE 713
QY 681 FYKRYKVLNRNLALPEVYRGKTSLLQYDASNSEMOGKTVPFLRESLEOKLEGRREE 740
DB 714 LYNNTKKTMPPDLARLDRLFCALFKALGLANENDYKGLTKVFRPGKPAEPDQIMSD 773
QY 741 VSHAAMVIRAHVGLFLARKQYKVLVC-----VYIIOKXVRAFLRR 783
DB 774 PDHLAELVK-RVNHMLTCSRWKQVOWCSLSVYKLKNKIKYBAEACIKQKTIHMLCKRR 832
QY 784 FLH-----LKAALVF-----OKOLRG-----QIARRYROL 811
DB 833 --HKPRIDGLVKGTLKRLDKFNEVSVLKDQKEMNKQIKNLEISIDTLMAKIKSTM 890
QY 812 AEKREOE-----KKQOE--EKKREBERERERRRREAB--LRAQ 851
DB 891 TQEQIOKEYDALVSSBELLSALQKKQOEBEAELRLIOEBMESGRKREDEGRKKE 950
QY 852 BEETKQOELALQSKQEAELITRELEKQENKQVEELIRLEKEITDLOMKKEQOELS 911
DB 951 BEERMKLEMEAKKQOEBERKREDDEKRIQAEVLAQLAKQE-----BESQQA 1003
QY 912 EASLOKQERRDQELRLLEEBCRAAQEFLSINDEFIDECVRIERSLSG----- 963
DB 1004 E-----OERRRREL-----ALRIQSEAEELIS---DEQAQDALRRNGSTRKMPTE 1047
QY 964 -----SFFSSE-----LAESACEEKNFNFQOPYEEZVDVDFEADD----- 1001
DB 1048 QMAKEMSEPLSGPAVLTKAAGTKKYDLSK-WKYAELRDTINTSCDIBLLACREBPH 1106
QY 1102 -----AFKDSPNPSEHGSDORTSGIRTSDDSESDPYMNDT 1038
DB 1107 RLKXYTHAMK-SKNKKNRTTTEQRAPKSVTDYDFA---PFLNNS 1146

RESULT 10
US-10-177-293-317
; Sequence 317, Application US/10177293
; Publication No. US200301024128A1
GENERAL INFORMATION:
APPLICANT: Lillie, James
APPLICANT: Glatz, Karen
APPLICANT: Zhao, Xumei
APPLICANT: Gannavarapu, Manjula
APPLICANT: Kamackar, Shubhangi
APPLICANT: Mertens, Maureen
APPLICANT: Myer, Vic
APPLICANT: Wang, Youzhen
APPLICANT: Xu, Yongyao
APPLICANT: Hoersch, Sebastian
APPLICANT: Monahan, John
APPLICANT: Meyers, Rachel E.
APPLICANT: Baet Jr., Robert C.
APPLICANT: Hortobagyi, Gabriel N.
APPLICANT: Puzstai, Lajos

```

APPLICANT: Meric, Funda
APPLICANT: Sahin, Aysegul
APPLICANT: Milla, Gordon B.
TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
TITLE OF INVENTION: PREVENTION, AND THERAPY OF BREAST CANCER
FILE REFERENCE: MRI-038
CURRENT APPLICATION NUMBER: US/10/177,293
CURRENT FILING DATE: 2002-06-21
PRIOR APPLICATION NUMBER: US 60/299,887
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 60/301,572
PRIOR FILING DATE: 2001-06-27
PRIOR APPLICATION NUMBER: US 60/306,501
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: US 60/325,002
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/362,585
PRIOR FILING DATE: 2002-03-05
PRIOR APPLICATION NUMBER: US 60/xxx,xxx
PRIOR FILING DATE: 2002-05-14
NUMBER OF SEQ ID NOS: 506
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 317
LENGTH: 1285
TYPE: PR
ORGANISM: Homo sapiens
US-10-177-293-317

Query Match 13.0%; Score 1391; DB 9; Length 1285;
Best Local Similarity 31.1%; Pred. No. 3,4e-72;
Matches 368; Conservative 225; Mismatches 365; Indels 226; Gaps 33;

10 LRENGHFPSTVSCAGIIVFRTDYGVFTYKOSTITQKVTAMHPTNEEGVDASLT 69
34 LNKGTFTLALIN-----QFPRAED-----SKDVEDCSLM 66
70 ELHGSIWYVLFQRYKRNQIWTYIGISLVNPPYQIAGLYEPATWEQYSRRHGLRPH 129
67 YLNEATLNLHNIKRYSGSDRIYTVANILAVNPYPIPKIYSEALKSYQSGSLGTRPH 126
130 IFPAINECTRLMKRDNOCILIKGSSGACKTESTLILKFLSVISQOSLESLKERTSC 189
127 VFPAIAKAPFDMVLYKMSQSIIVSGSGAGKTEITFVLRYLTESYGTQD----- 177
190 VERALIESPIMAFNGAKTVVNNSSRFQVQNLNCKGNIOGRIYDCILSSONRVV 249
178 IDRIYEAHPLEAFENAKTVRNNSRRGKFEIHPNEKSSVVG-PSHYILEKSRIC 236
250 RQNGERNYHIFYALLAGLEHEEREEFYLTENNYHYN----- 288
237 VQGEKERNYHIFYRLCAGASEDIRKHLSSPDNFRYLRNGCTRYFANKETDKQILQNRK 296
289 -----OSGCVEDTIDDSREFVITAMDMQFSKEVEVSLLAGLILHGNIEF----- 339
297 SPEYLAGSKWDLDDHGFIMCTAMKKGIDDEKDLDFRVVAGVHLGIDDEEAG 356
340 ITAGAGVSPKTA--LGRSAELGLDPTQTLDTLQJRSF-----LRGEELILPLVVOO 392
357 STSGCGNLKXKKSQSLFYCAELGLDODLRLVSLTRWMLTTGRTGKVIKPLVEQDA 416
393 VDSRDSLAMALVACCFEWMYIKKINSRIKGNEDKSGIGLIDFGENFEVNHPEQFINYA 452
417 NNARDAALAKTVYSHLPDHVYVNRNQCPEFTSSYFGLVDIAFEYFEHNSPFCINC 476
453 NEKLGQYFNKHIISLEQLEISREGLYWEDIDWIDNECIDLTKU--GLLALINESHP 511
477 NEKLGQYFNKHIISLEQLEISREGLYWEDIDWIDNECIDLTKU--GLLALINESHP 536
512 QATDSTLLEKLSCHANNHFFV-----KPRVAVN-----NFGVYAGBVOYDVAGIIE 560
537 QSDQDHTAVNOKH-KDHFRLTIPKSKLAVVRNIRNDBEGFTIRHPAGVACTETTPAE 595
561 KNDTRFDLLNLRESRDFIYDLFEHVSRRNODTLKCGSKGRPRPVSSQFVDSLHS 620

596 KNDALHMSLESLICESBDKFIREFES-STNNNDTKQKAGKSLFISVGNKFKT-QNLT 653
621 LMATSSSNPFRVRIKXNMQMPQPDQAVVLNOLRSGSLVETVIRIKAGVAVRPPQD 680
654 LIDKLRSTGASTIRCIKIKNLKWTSHHFEQAOLISQCSGWSVLDLQGGPSPASFHE 713
681 FYKRYVLMRNALBEDVYRGKTSILQYDASNSSEWQGLTKVFLRESLEQLEKREBE 740
714 LYNNYKXMPDLARLDRLPCKALFKALGNENYKGLTKVFPKPKAFEPDQIMSD 773
741 VSHAAMVIRAHVGLFARKQYRKVLVC-----VVITQKYRAFLLRR 783
774 PDHLAELVK-RVNHMLTCSRWKQVWCSTLSVYKLNKIKYPAEACIKMQKTIKRWLCRR 832
784 FLH-----LKKAAVLF-----QKQARG-----QIARVYRQL 811
833 --HKPRIDGLVKVGLTKRLDKFNEVSVLKDGRPNKQIKNLEISIDTLMAKIKSTM 890
812 ASKREQEE-----KKQEE--EKKKREERERERERE--LRAQ 851
891 TQEQIQKQYDALVKSSEELLSALQKKQEEBAELRIQEMEKERRREDEKRRRE 950
852 EETRKQOELEALQSQKBAELTRELEKQENKQYEBILREKTEIDYQRMKEQOELSLT 911
951 EERERMKLEMEAKRQEEERKREDEKRIQAEVLAQLOKE-----EESQQAVAL 1003
912 EASLOKQERPOQELRLBEACRAQFELSINDDELDECVRNTERLSG----- 963
1004 E-----BERRREL-----ALRTQSEBAELIS--DEQADALARNQDTRPKMTP 1047
964 -----SEFSSE-----LAESACEEKNPNFQOPYEEBVDSGEFADD----- 1001
1048 QWAKEMSEFLSGPAVLATKAAGTKYVDLSK-WYALRLDTINSCIELLAACREBH 1106
1002 -----AFKOSPNSBEHSHSDQRTSGIRTSDDSEEDPYWNT 1038
1107 RLKTVYHAWK-SKNKKRTEQORAPKSVTDYDFA---PFLNNS 1146

RESULT 11
US-10-044-303-2
Sequence 2, Application US/10044303
Patent No. US20020137161A1
GENERAL INFORMATION:
APPLICANT: Max-Planck-Gesellschaft e.V.
TITLE OF INVENTION: Protein expression and structure solution using
FILE REFERENCE: ST010209-EPA
CURRENT APPLICATION NUMBER: US/10/044,303
CURRENT FILING DATE: 2002-01-11
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 1016
TYPE: PR
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: Whole sequence
OTHER INFORMATION: The protein comprises as component (1) the aa
OTHER INFORMATION: Sequence of myosin II motor domain of
OTHER INFORMATION: Dictyostelium, a three aa linker region and the
OTHER INFORMATION: a-actinin aa sequence
US-10-044-303-2

Query Match 12.7%; Score 1360; DB 12; Length 1016;
Best Local Similarity 34.3%; Pred. No. 1.6e-70;
Matches 354; Conservative 207; Mismatches 360; Indels 112; Gaps 29;

8 VLRENGHFPSTVSCAGIIVFRTDYGVFTYKOSTITQKVTAMHPTNEEGVDASLT 62
39 IWYNPD-----PKERDYSYCGEIVSET--SDSFTFTVVDGQDQVKKDDANQRNPIKFDGV 92


```

Qy 591 -----SRNODTLKCGSKHRR---PTVSSQFKVDSLSHSLMATLSSSPFVACIKENMG 641
Db 626 VGLDQAKAMTESSLPSSAKTKKGMFRVGOGLYK-EQGLKMTLRLNTTPFVACIIPNHE 684
Qy 642 KMPDQFOAVVLAOLRYSGMLFVTRIRKAGVAVRPPDFYKRYKVLMMNLALPE---DV 698
Db 685 KRSGKIDAPLVLEQRLCNGVLEBGRICRQGFPMRIVQFEPFRORYEILAAAN-AIPKGFMDG 743
Qy 699 RGKCTSLLOLYDASNSEMOLGKTRVFLRESLBOKEKREEREEESHAMVIRAVLPIPLAR 758
Db 744 KQACILMIKALBEIDPNLIRIGOSKIFRFTGVLAHLEBERDLKTTDVIMAFQACRGYLAR 803
Qy 759 KOYRK---VLYCVIIOKNYRAFLRRFLHKKAAIVFOKOLRGQIARRVROL--LAE 813
Db 804 KAPAKQOQLTANKVIOGNCAVYL-----KLRNOMWMLPFKVPVPLQ 846
Qy 814 KRQOEKQOEKKEEKKREBEREREREREAELRAQO--EETTRQOQLEALQSKQKA 871
Db 847 VTRQEBEMQAKEDLOKTKERQOKAENELKELOKHSQLEEFKULLQEOLOA-----ET 900
Qy 872 ELTRELEKOK-----ENKOVEEIL-----RLEKEIDLOMKQOELSLTEASLOKLOER 921
Db 901 ELTAEBEBKVRILAACKQOELBEILHEMEALHEEBEDRGQOLOAKERK-----KMAQOK 952
Qy 922 RDOELRLREBACRAAOEFLESINPDEI--DECVRIERSLSGSEFSSELAESACEKXP 979
Db 953 LDLEBOQEEREAAR-----QXLOLEKVTAEKIKYLE-----984
Qy 980 NMFPSQPYREBEVDEGEADDDAFKOSPNSBHGSDORTSGRTDSDSEEDPYANDTV 1039
Db 985 -----DEILVDDONNKL--KERKLEBRISDL-TTNLAEBEERKAKNLTJK 1027
Qy 1040 VPTSPSADSTVLLAPSVQDSGLHNSSGESTYCPMPONAGDLPSPGDYDODDDYEDGA 1099
Db 1028 LKRNK-----HESMISELEVRLLKE-----EKSKOELE---1054
Qy 1100 ITSGSVTSNSYSGQSPDYRCSVGTYSNGSAYRFSSEGAQSSPFEDSEEDPDSRF-DYD 1158
Db 1055 -----KLRKLESDADDFHQIADLQAOJAEIK 1082
Qy 1159 DELSTYRDSYSCVTLPYFHSFLYMGGLMNSWKRMVCVLDFTFLMPSKQALQGMW 1218
Db 1083 MQLAKKEEBLOALAR-----LDDE-----IAOKNNALKK--1 1113
Qy 1219 HKKGGSSSTLSRNWKKRMFVLROSKLUMFENDSEKLGTVVTRAKELIONTTKENG 1278
Db 1114 RELBGHISDLOEDLSEBAARKAERKOKDGLSELEALKTELE---DPLDSTATQO-- 1166
Qy 1279 DIIMADRTPLHLESBEDASQMFVLSOVHASTDOEIOEMHDEQANPONAVGLTDLGLID 1338
Db 1167 -----ELRAKREGEVTLKKAALDEBTRSHQAQVQEMROHQAQ-----1203
Qy 1339 SVGASDPSDBPNSFVITIANRVLHCAADTPREEMHMTITLLQNSKDPTRYEGGEFIVRGW 1398
Db 1204 --AVEBELTQLEQFKRAKAN--LDKMKQOTLEKENADLA-----GELRVLGQ-----A 1246
Qy 1399 HKEVKNSPKMSLSLKKRMFVLTHNSGLDYKXSEKXALKGLTVLNSLSGVAPPEDEKIRK 1458
Db 1247 KQVEHKKKKLEAQVE-----LOSCKS--DGERARA 1276
Qy 1459 ETGYMNVTVYGRKHCRYLTKLNEA-----TRMSVITQNVTTDKAP 1500
Db 1277 EL---NDKYHKLQNEVESVTGMLNEAEGKAIKLAKOVASISLOLOTOGELBOETROKAN 1333
Qy 1501 IDTPTOOLLODIKENCCLNSDVVEQIYKRNPIRLYTHPHLPLPLPYGDININLIXDG 1560
Db 1334 VSTFKRLQLEEBE--RNSLOLODLEMEAKONLBRH-----ISTLNIQLSDSK- 1377
Qy 1561 YTTLOD-----EAI-----KIFNSLOQLESMSDPIPIIGILQTOHDLRPLRDELXY 1606
Db 1378 -KTLQDPASTVALEBGGKRFQKEIENLTQOYBEKAAAYDKL--KTYNRLOQOELDLIV 1433

```

```

Qy 1607 COLIKOTNKVPHPGSVGNLYSMOILTCLSCFPLPSRGILIKYLKRIREOPGTEMEX 1666
Db 1434 VLDLNRQOLVSNLEKKORFQ-QLLA-----EKNVSS--KYADERDRAAEAREKET 1483
Qy 1667 YALFYTESLK-----KTKREFPVSRDEIEALIR-----1696
Db 1484 KALSLARALEALKEAKEBELERTNKMLKAMEDLVSSKQDVGNVHLEKSKRALBETOMBE 1543
Qy 1697 -----QEMSTVYCHGGGCKTITNSHTTAGEVVEKLIRGLAMEDSRN-----MFLF 1744
Db 1544 MTOLEBELEDELOATEDAKLRLEVMOALKGO----FERDQARBQNEBKROLOROLH 1599
Qy 1745 EYNGHVDAKIESRTVADVLAKFEKLAATSEVGDLPWKYFPLCYCLDTPNVKOSVEPA 1804
Db 1600 EYTELEDERKQORALAAAKKLE-----GDL-----KO-----1628
Qy 1805 PMFEOAHEAVITHGHHPAPENIOVLAALRYLOQDYLTHAAIPPLEVYYSIQR-----1858
Db 1629 --LELOADSAIKGR---EBAIKOLRKLQAOQKDFORELEBDARASRDEIFATAKENEKKA 1682
Qy 1859 --LKARISQSTKTFPPCER-----LEKRR-----TSFLEGLTRRSFRFGSVYR 1899
Db 1683 KSLBADLMQLOEDLAAEBARQADLEKEBELAEBLASSLSGRNALQDEKRRLEARIQDE 1742
Qy 1900 QKVEBEO-----MLDMWK-----EEVSSARASIIDKRRKFGQMNQ--EQAMAKYMALIK 1947
Db 1743 BELEBEOGMEMMSRVRKATQAOQLENSLATERSTQKNSANQOJEBQNKELRSKLH 1802
Qy 1948 EWPQ-----YGSTLPDVECK 1962
Db 1803 EMEGAVKSKFKSTIALEAK 1822

```

RESULT 13

```

US-10-171-311-162
; Sequence 162, Application US/10171311
; Publication No. US20030087270A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Chen, Yan
; APPLICANT: Zhao, Xumei
; APPLICANT: Monahan, John
; APPLICANT: Kamatekar, Shubhangi
; APPLICANT: Glatt, Karen
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Hoersch, Sebastian
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; FILE REFERENCE: MRI-035
; CURRENT APPLICATION NUMBER: US/10/171,311
; PRIOR FILING DATE: 2002-06-12
; PRIOR APPLICATION NUMBER: US 60/298,159
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,155
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/335,936
; PRIOR FILING DATE: 2001-11-14
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162
; LENGTH: 1972
; TYPE: PRF
; ORGANISM: Homo sapiens
US-10-171-311-162

```

Query Match

12.4%; Score 1333; DB 9; Length 1972;

Best Local Similarity 24.2%; Pred. No. 1.4e-68;

Matches 514; Conservative 362; Mismatches 746; Indels 498; Gaps 72;

```

Qy 8 VWLRNQGHH--PSTVNSCAGIYVFTDYQCVFTYKQSTTTHQCVTAMHPTNEGVUDM 65
Db 36 VVVPSEKQGFPAASIKERKGDVENVELVENG-----KRVTVGKODIQGMNPKFSGKVEDM 90

```

Qy 66 ASLTJLHGGSIWYNLFQRYKRNQIWTYIGSILASVNPYOPJAGLYEPATMEQYSRRHJGE 125
 Db 91 AELTJCNESAVLHNRERYFSGILITYSGLFQVNVNPGHLP-ISEKELVDVYKKGKRE 149
 Qy 126 LRPPIPALINECYRCLMKRHDNOCILIKESGAGKTESKILKPLSVISQOSLESLKE 185
 Db 150 MPPHYALADTYRMLDREDOISLCTGESGAGKENTKVIQYLAIVASSHKKGDTS 209
 Qy 186 KTSCEBALLESPPIMEAFNAKTYVNNNSRFGKVOLNICOKNIOGRIIVDCILSSQ 245
 Db 210 ITGELKQOLQANPILFENAKTYKNDSSRFGKIRINFDVTGYIVANI-ETYLEK 268
 Qy 246 NRVNQNPGERNYHIFYALLAGLEHREBEFYLPSTPENYLYNOSGYEDKTIISQESFR 305
 Db 269 SRAIQARDERTFHTFYWIAKAKEMRSDLLBEGNNYTPLS-NGFVPIPAQODDMQ 327
 Qy 306 EYITMDVMOFSKEKEVREVSLLAGILHGNIEF--ITAGAQVSFKTALGRSABLLGL 362
 Db 338 ETVEMAIMGFSEEBOLSLKYVSSVLOIGNIVFKKERNTDQASMPDNTAAQKCHLMGI 387
 Qy 363 DPTOLT-DALTOBSMFLGEEILTPLVQOAVD-SRDSLAMALYACCEFWIKKINSRI- 419
 Db 368 NITDTRSLTPRIKXV--GRDVQKQKQTKQADPAVEALAKATYERLFRWILTRVAKAD 445
 Qy 420 -KGNEDFKSIGILDIGFENFEVNFQEPNINYANEKLOHYENKHI FSLIOLEYSGEIV 478
 Db 446 KTHRGASFLGIDLAGFEIFEVNSPEQICINYTEKQQLFNHMFLEQEBYQREGIE 505
 Qy 479 WEDIMW-IDNGECDLIEK--KGLIALINESSHPOATDSTLEBKHSQHANNHFFYK 534
 Db 506 WAFIFGDLQPCIEILIERPNPFGVALLDEECWFPKATDKSPFEKCTEGESHKPFQK 565
 Qy 535 PRVAVN--NFGVHYAGEVOYDVRGILEKNRDPFDLNLRESRPFIDYLFPHVS- 590
 Db 566 FQOLDKTEFSIINHAGKVDYNASAMLTQMPDLNDVTSILNASSDKFVADLMKQVDR 625
 Qy 591 -----SRNODTLKCGSKRR--PTVSSQPKVDSLHSJMATLSSNPFPRCIIKPMQ 641
 Db 626 VGLDMQAKTESLPSASTKKGMPRTVQLYK-BQLGKMTTLNTTNPFRCIIPNE 684
 Qy 642 KMPDPDQAVNLNOLRYSGMLETVRIRKAGIVRRPPOD FYRYKYKMLNLAPE--DV 698
 Db 685 KRSGLDIAFVLEQRCNGVLEGIRICRQGFNRIVFOE FRQRYEILAN-AIPKGFMDG 743
 Qy 699 RGKCTSLIQLVDASNSEWQKTYVFLRESLEQKLEKRESEVSHAAMVIRAHVGTAR 758
 Db 744 KOACILMTALELDPVLYRIGOSKIPFTGVLAHLEBERDLKITDVIMAFQAMCRGYLAR 803
 Qy 759 KOYRK---VLYCVIIOKNYRAFLRLRRFLHLKKAIVFQQLRQIARVROL-LAE 813
 Db 804 KAFARQOQLTMKVIQRYCAVY-----KLRWQWQRLTLTYKXPLIQ 846
 Qy 814 KREQEKKKQEBEKKKREBERERERREALBAQO--EETTRKOQELALQSKQEA 871
 Db 847 VTRQEEHQAKEDLOKTEROQKAKENLEKEQKHSQLTEKKNLLOEQLOA-----ET 900
 Qy 872 ELTRLEBKQ-----ENKQVEEIL-----RLKEIJDLORMKEQOELSTESLOKQER 921
 Db 901 ELYABAEERKRVLAANKQOELIHLHMEARLEEBEDRGQLOAERK-----KQAQOM 952
 Qy 922 RDOELRLLEEBACRAOEFLESINPDEI--DECVNIERSLSGSEFSSSELAACEKP 979
 Db 953 LDLEQGLEEBEAR-----QKLOLEKVTAAEKIKLE----- 984
 Qy 980 NFNFSQPIPEEVDGCFEADDAFKDQPNPSHGSHDQFTSGIRTSDDSEDPYMNADV 1039
 Db 985 -----DLTVMDDONNKLS--KERKULEERISDL-TTNLAEEBKAKANLTK 1027
 Qy 1040 VPTSPASDSTVLAPSVQDSGLHNSSGSESTYCPQNGDLPSPDGDVDVYQDDYEDGA 1099
 Db 1028 LKNK-----HSMISELEVRLLKG-----EKSRQELE--- 1054

Qy 1100 ITSGSSVTFSNSYGSQWSPDYRCVGTYNSSGAYRPSSEGAQSSFEDEEDFDSRF-DTD 1158
 Db 1055 -----KLRKLEGGASDFHEQIADLOAQIABLK 1082
 Qy 1159 DELSVRDSVYSCVTLPIFHSFLYKGGSLMWSKRWKCVLQDETFLMFRSQOALQOML 1218
 Db 1083 MQLAKKEBELQALALR-----LDDE-----IAQRNALKK--I 1113
 Qy 1219 HKGGGSSTLRRRWKRWKFWLROSKLWYFENDSEBKGTVEVYTAKEIINDTKENGI 1278
 Db 1114 RELEGHIDLDDBDSEBAANKAEKQKRDGELEBALKTLE-----DTLDDSTAQ-- 1166
 Qy 1279 DIWADRTFHLIABSPEDASQWFSVLQVNASTDQIOEWHDEQANFQNAVGLDVLID 1338
 Db 1167 -----ELPAREQGVTVLKALDEBTRSHBAQVEMQKHAQ----- 1203
 Qy 1339 SVCSADSDPRNSPFIITANVNLHGNADTPREMHMITLORSKQDTRVBOGFVRCWL 1398
 Db 1204 --AVELETOLEQFRAKAN--LDKNKOTLEKENADLA-----GELRVLQO--A 1246
 Qy 1399 HKEVNSPMSLSLKUKKRWFLTHNSLDYKSSSEKNALKGLTVLNSLCSVVPDEKIFK 1458
 Db 1247 KOEVEHKKKLEAOVQE-----LOSKCS--DGERARA 1276
 Qy 1459 ETGYMNVTYGRKCYRBLTYKLNEA-----TRWSSVIONVTDTRAP 1500
 Db 1277 EL--NDKVHLQONEVESVLTGMLNABEGKAIKLAQVASSLSQLODTELLQEBETROKLN 1333
 Qy 1501 IDTPQOQIIOIKENCLNSDVVEQIYKRNPLRTYTHNPLHPLPLPYGGINLNLKKG 1560
 Db 1334 VSTKRLQLEER--RNSLDQDDEEBAQONERH-----ISTINIQLSDSK- 1377
 Qy 1561 YTLLOD-----EAI-----KIFNSLOLESMSDPIPIIOGILQTHDRLPRJDEY 1606
 Db 1378 -KLODPASTYBALLEGGKRPQKEITENULTQOYEEAAYDLE--KTRAKLOEGLDV 1433
 Qy 1607 QOLIKQTNKVPHPQSVGNLYSMOILYTCLSCFPLSRGILKYKFLKRIREQPGETEMK 1666
 Db 1434 VDLNQOQLVSNLEKKQKFD-QLLA-----BEKNIS--KVADBRDRAABAREKET 1483
 Qy 1667 YALPTYESLK-----KTKCEPFPSPROEITALLHR----- 1696
 Db 1484 KALSLARLLEBALBEKELBERTNKULKAEMEDLVSSKODGVKNHLEKSKRALETOMBE 1543
 Qy 1697 -----QEMTSVYHCGGSSCKITINSHTTAGEVVEKILRGLAMEDSN-----MPALF 1744
 Db 1544 MKTQLELEDELOATEDAKLLEVNMOLKQO---FERDLOARDEJNEBRQLOQROLH 1599
 Qy 1745 EYNGHVDKALIESRTVAVDVLAKFEKLAATSEVGDLPWKFYFKLYCFLDTDVVPKDSVEFA 1804
 Db 1660 EYETLEDEBERQORALAAAKKLE-----GDL-----KD----- 1628
 Qy 1805 FMFOAHBAVJHGHHPABENULQVLAALRLOYLOGDYTLHAALPPLBEVYSLOF----- 1858
 Db 1629 --LELOADSAIKGR---EBAIKOLRLQAOQKDFORELEBARARSROIPTAKENEXKA 1682
 Qy 1859 --LKARSQSTKTFPPCR-----LEKRR-----TSFLEGLTRBSFRGSVVR 1899
 Db 1693 KSLLEDLQLOEDLAAARAKQADLEKEBLAELASSLGRNALQDEKRRLEBARIAQLE 1742
 Qy 1900 OKVEEQ-----MLDMWIK-----EEVSSARASIIDKRRKFOQNMQ--EQAMAKYALIK 1947
 Db 1743 EELBEEOGMEMASDRYKATQOQLOLSNELATERSSTQKXBSARQOLERQNKELRSKLN 1802
 Qy 1948 EWPQ-----YGSTLPDVECK 1962
 Db 1803 EMEGAVYSKFKSTIAALEAK 1822

RESULT 14
 US-10-044-303-1
 ; Sequence 1, Application US/10044303
 ; Patent No. US20020137161A1

/ GENERAL INFORMATION:
/ APPLICANT: Max-Planck-Gesellschaft e.V.
/ TITLE OF INVENTION: Protein expression and structure solution using
/ FILE OF INVENTION: Specific fusion vectors
/ FILE REFERENCE: ST010209-EPA
/ CURRENT APPLICATION NUMBER: US/10/044,303
/ NUMBER OF SEQ ID NOS: 3
/ SOFTWARE: Patentin Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 765
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: Partial myosin
/ OTHER INFORMATION: sequence of Dictyostelium; Component (1) of the
/ OTHER INFORMATION: recombinant protein M761-2R R238E
US-10-044-303-1

Query Match 12.4%; Score 1329; DB 12; Length 765;
Best Local Similarity 40.1%; Pred. No. 6; 8e-69;
Matches 299; Conservative 160; Mismatches 253; Indels 34; Gaps 16;

QY 8 VMLRENGQHPSTVNSCAGIIVFRTDYGQVFTYK-----QSTTHQKVTAMHPTNEEGV 62
DB 39 IWNPD-----PKERDYECEGELVSEY--SDSFFFTVDGQDRQVKDANORNPFIKFDGV 92
QY 63 DDMASLTSLHGGSIWNLFORYKRNQIMWTYGISLASVNPQPIAGIYEPATMEQYSRRH 122
DB 93 EDMSLSYINBPVAFNLRVYNQDILYISGLFLVAVNPFKIP-ITYQEMDPIFKGR 151
QY 123 LGLPPIFAIANECYRCMLKRDNOCILIGESGAGKTESTLILKFLSVISQSLBLS 182
DB 152 RHEVAHIFAIISVAYRSMIDRQNSLITGESGAGKTEMTKVIVYLAAYAGR-----N 207
QY 183 LKEKTSCEVERALIESPIMEAFGNATVYNNNSRFGKVQVQNLICQKNIQGGRIYDCL 242
DB 208 QANGSVLLEQOILQANPILFAFNATKTRNNNSRFGKFIETIOFNSAGFISGASIOSYLL 267
QY 243 SSQNRVVRONGERNYHIFYALLAGLHEBEREEFYLTSPENHYLNQSCVEDKTI SDPE 302
DB 268 -EKSRRVVFQSETERNHYIYOLAGTAEBEKKALHAGPESFYLNQSCVEDKTI SDPE 326
QY 303 SFREVITANDVQFSKEEVRERSLAGILHLGNIEFIT-AG-GAQSFKTALGRSABL 360
DB 327 EFKITRQAMDIVGFSQEGQMSIFKTIAGILHLGNIEFKGAGGAVLXKXTLMASTYF 386
QY 361 GLDPTQLTALTRQSMFLGEBELTLPNQAQVDSQSLAMALYACCFEWTYKINSRIK 420
DB 387 GNPSTLEKALMPRIAGRDVAQHLNVEKSSSSRDALVKALYGRFLPMLVYKINNVLC 446
QY 421 GNEDFSIGILDI FGFEFENFVNHFEOFNINYANEKIOEYFNKRIIFSLDELEYREGLVWE 480
DB 447 QERKAVFTGLVDSGEIFRQNSFEQLCINYNTEKIOEYFNHMFLEDEEYLUKEKINT 506
QY 481 DIDW-IDNGECLDLIEKUL--GLLALINEESHFPQATDSTLLEKLSQHA--NNHFYVQR 536
DB 507 FIDFGDSQATIDLDIGRQPGIALLDQSVFNPNTDNTLITKLSHRSKQAKKEER 566
QY 537 VAVNNGVYHAYAGEVOYDVAGILEKRDTPFRDDLNLRLBSRFDIYDLFE--HSSRN 594
DB 567 FSKTEGVTYHAGQVWEYIOWMLEKNKDPLOQDLIECFQDSSDNVVTKLFPNPINIASR-- 624
QY 595 QPTLKSGHRRFTVSOPKVDLSHLMATISSNFFVRCIKPNQKQKPDQDAVVLN 654
DB 625 ---AKGANP--ITVAQYK-EQLASLMATLSTTNHFRCLIPNNKQLPALKEDEVRVD 678
QY 655 QURSGMLETVIRIRKAGVAVRFPQFYRYKYVLMENLAL-PEDVAGKTSLLQUTDASN 713
DB 679 QURNGVLEGIIRTRRGFPRIIYADFVIRYILANVPRDAEDSKATDAVLKHLNIDP 738
QY 714 SEMQAGKTVFLRESLEQLEKRRER 739
::: |||:: |||::

DB 739 EQYREGITKIFFRAGQGLARIEAREQ 764

RESULT 15

US-09-927-597-2
/ Sequence 2, Application US/09927597
/ Publication No. US20030032018A1

/ GENERAL INFORMATION:
/ APPLICANT: Malik, Fady
/ APPLICANT: Beraud, Christophe
/ APPLICANT: Freedman, Richard
/ APPLICANT: Craven, Andrew
/ APPLICANT: Sakowicz, Roman
/ APPLICANT: Hartman, James
/ TITLE OF INVENTION: Human smooth muscle myosin heavy chain
/ FILE REFERENCE: CYTOPO18
/ CURRENT APPLICATION NUMBER: US/09/927,597
/ CURRENT FILING DATE: 2001-08-10
/ NUMBER OF SEQ ID NOS: 17
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 2
/ LENGTH: 1945
/ TYPE: PRT
/ ORGANISM: Human
US-09-927-597-2

Query Match 12.4%; Score 1328.5; DB 9; Length 1945;
Best Local Similarity 24.2%; Pred. No. 2; 4e-68;
Matches 515; Conservative 365; Mismatches 742; Indels 505; Gaps 74;

QY 8 VMLRENGQHP--PSTVNSCAGIIVFRTDYGQVFTYKOSTTHQKVTAMHPTNEEGVDM 65
DB 36 VWPSEKQGFENASIKERKGEVVELVENG-----KKTVGKODIQKNNPFPKSKVBDM 90
QY 66 ASLTSLHGGSIWNLFORYKRNQIMWTYGISLASVNPQPIAGIYEPATMEQYSRRHIG 125
DB 91 AELTLNBSAYLHNREYFSGLIYTGCLCVVNPYKHLR-ITYSEKIVDYKGRKHE 149
QY 126 LPPHIFAIANECYRCMLKRDNOCILIGESGAGKTESTLILKFLSVISQO---SLELS 182
DB 150 MPPHIALADNAYRSMIDRQNSLITCTGESGAGKTEMTKVIVYLAAYASHKQKOTS 209
QY 183 LKEKTSCE---VERALIESPIMEAFGNATVYNNNSRFGKFIETIOFNSAGFISGASIOSYLL 238
DB 210 ITQGSFAYGELKQLQANPILFAFNATKTRNNNSRFGKFIETIOFNSAGFISGASIOSYLL 268
QY 229 DCILSSQNRVVRONGERNYHIFYALLAGLHEBEREEFYLTSPENHYLNQSCVEDKTI 298
DB 269 EYTLLEKSRARQARDERTFHI FYMIGAKRKMSDLLLEGFNNTYFLS-NGFVPIPA 327
QY 299 SPOESFREVITANDVQFSKEEVRERSLAGILHLGNIEF--ITAGGAVSFPTALGR 355
DB 328 QDDMFQETVEAMALMGSEEBQSLIKLVSSVLDGNIIVFKKENTQOASRPDNTAAQ 387
QY 356 SAEILGLDPTQLT-DALTQRSNMFRLGEBELTLPNQAQVDSQSLAMALYACCFEWTYK 413
DB 388 VCHLGINVITPSTRILTPRIKV--GRDVVQAQKTEQADPAVEALATATYERLERWILT 445
QY 414 KINSRI--KGNEDFSIGILDI FGFEFENFVNHFEOFNINYANEKIOEYFNKRIIFSLDE 471
DB 446 RNNKALDTRHGRGASFLIDLAGIEFIVNSFEDLCINYNTEKIOEYFNHMFLEDEE 505
QY 472 YSRBGLVMEIDW-IDNGECLDLIEK--KGLLALINEESHFPQATDSTLLEKLSQHA 527
DB 506 YOREGIEWNPIDFGDLOPCTIELIRPNPFGVALLDBECMFPAATKSPYEKICTQOG 565
QY 528 NNHFYVQRVAVN--NFGVAYHAYAGEVOYDVAGILEKRDTPFRDDLNLRLBSRFDIYDL 585
DB 566 SHPKQKQKQKDKTEFSGIHYAGVDYNASAWLTKMNDPLNDNTSLNASSDQKFAVDL 625
QY 586 FEHVS-----SKNNDPTLKSGSKRR--PTYSQPFKVDLSHLMATISSNFFVRC 634
DB 626 WQDVIRIVLDQMAKMTESLPSASKTKKGMFRIVGOLYK-EQLKLMTTLNTPNPFVR 684
::: |||:: |||::

Qy	635	CI KPMOMKMPDFOADVUNLONLRYSGMYETVYIRKAGYVVRPRPDOPFYKUYKVLNRYNL	691
Db	665	CIIPNHEKSGKDLFVLEOJRCNGVLEGIPI SCQGFNRRVLFVQFRRYETELIAN -AI	743
Qy	695	PE---DVAGKTSYLSQLYDASNSEWOQKTKYFLEBSLEBOUKLEKREBEVSHAAMVIRAH	751
Db	744	PKGFMDGHOACIMLKALELDPNLVIRIGSKIFPRIGVLNHEEBDLKITDVIMAFQM	803
Qy	752	VLGFLARQOYK---VLYCVIIOKNYAFLLRRFLHKKALYIFOKOLRQOIARVYR	808
Db	804	CRGYIARAKAFARQOOLTAMKYIQRCAVY-----KLRMWMWRRLPT	846
Qy	809	OL--LAEKROBEKKKEEKKKEEREREREREREAELRAOQ--BEETRKOELATL	864
Db	847	KYRPLQYTRQOBEWQAKEDLOKTKERQOKALENEKLEJOCHSOLTEKKNILQEBLOA-	905
Qy	865	OKSQKEALITRELEKOK---ENKOVEITL-----FLEKTEIDLOKMEQOELSTEAS	914
Db	906	-----ETELVYAEEMRVLAKKOBLEIITHEMEARLEEBERDGOLOAKER-----	953
Qy	915	LOKUORRDOELRLREESACRAOELBSINDEI--DECVANIERSLSGSEFSEELAE	972
Db	954	-KMAQMDLBLEQLEEBAR-----QLOLEKTAEKAKIKYLE-----	991
Qy	973	SACEKPNFNFSPQPYREEVDEGFREADDAFYDSPYSECHSDORTSGIRTSDDSEED	1031
Db	992	-----DELVMDQNNKLS--KBRKLLBERISDL-TTNLAEBEE	102
Qy	1033	PYMDTVVPTSPSADSTVILLASVODSGSLNNSGSESTYCMPONAGDLPSPGDYDYO	109
Db	1028	KAKNLTJTKNK-----HESMISELVEVRKKE-----EKSR	105
Qy	1093	DVEDGATSGSVTFPSNYSQSWSPDYKCSVGTYNSSGAVYPPSSGAQSPFDESEED	115
Db	1058	QELE-----KAKRLEBGASDHEQIADQ	108
Qy	1153	SRF-DTDELSTYRDSVYSCVTLPYPHSFLYMKGLMNSWKRCVYLKDETLMFRSKOE	121
Db	1083	AOIHELKQOLAKKEEELQALMAR-----LDDE-----IAQKN	111
Qy	1212	ALKGOMLKKGGSGSTSLSRMKKRRKMFVLROSKLMFENDSEKKGTEVATKEIN	127
Db	1116	ALKK--TELBSHISDLOEBDLSERAKNKAEBQOKRLOEELBALTYELE---DTLOS	116
Qy	1272	TTKENGIDDIIMADRTFHLIAESPREDSQMFVLSQVNASTDDEIOEMHDEQANPONAVGT	133
Db	1169	TATQO-----ELRAKREGEVTVLKYALDEBTRSHNAQVOEMRQKNQ-----	121
Qy	1332	LDVGLIDVCSASDSDRPNRSPYITTAANRVLHCNADTPREMNIMITLLQSKDPTREVOE	139
Db	1211	-----AVEELTEOLOEFKRAKAN--LDKNQOTLEKENDLA-----GELRYLGO-	125
Qy	1392	FIVRGMLKHEVNGSPYMSLSLKKKKMFVLTNHSLDYUKSEKNALKLGTVLYNLSLCSVVP	145
Db	1253	-----AKOEVENKKKKLEAOVOE-----LOSKS---	127
Qy	1452	PDEKIFKETGWNVTYUGRKHCVYLYTKLLNEA-----TRWSSVION	149
Db	1277	DGERARAEI---NDKVNKLQNEVESUTGMLNABEGKAIKLAQVAVSLSQLODPOELLOE	133
Qy	1494	VJDTAARIDTPROOLIOIKENCSLSDVVEOYUKNPIIARYNHPHLSFLRPLYUDITL	155
Db	1334	ETROKLANSTKIKROLEE--RNSLDQODDEEMAKONLERRH-----ISTLNI	137
Qy	1554	NLKDQGYTTLOD-----EAI-----KTFNSLOQESMSDPIPIQIGILQIOTGNDH	159
Db	1379	QUSDKK--KKQODPASTVEALBEGKKRFOKEIENLLOUEBKAAVADKLE---KITNRQ	143
Qy	1600	PLRDELVCOLLQOTNKVPHPSVGMLYSQOILTCLSTFLPSBGILIKLYKFLHKIRIEOF	165
Db	1434	QELDDLVVDLDNOROLVSNLEKKORFQD-OLA-----EENKUNSS--KYUDERORAA	148

```

0Y      1660  PGTEMEXKALPTYVSLK-----KTKCRFPVSPSDEI:EA:LI:HR-----1696
           |||
           :|:
           :|:
Db      1484  EAREKETKALSLARALEEALAEKELERTNKMLKAEMEDLVSS:CDVDGKXVHLE:EKSKRA 15433
           |||
           :|:
0Y      1697  -----QEMTSTVYCHGGGSCKITINSHTTAGVEVWEKLI:RGLAMEDSRN-----1739
           |||
           :|:
Db      1544  LETQMEEMKQTLEBELEDELQATDEDAKURLFLENNQALGQ-----PERULQARDEQNEBEKRR 15939
           |||
           :|:
0Y      1740  --MALPEYNGHVDKALIESRTVADVLAKPEKKA:LAATSEVGDL:PMKFFYKLYCFLDTONVP 17977
           |||
           :|:
Db      1600  QLOQOLHEYELEFELDERKORALAAAKKLE-----GDL-----16333
           |||
           :|:
0Y      1798  KDSVEFAPMFQAEVAVIHHHRAPEENLOALALRYLOV:LGQDVTYLAHAI:PLPEVYSLO 18577
           |||
           :|:
Db      1634  KD-----LELOQDSAIKGR---EAAIKOLRLQLOQMDFORELEDA:RARSDEIFATA 16822
           |||
           :|:
0Y      1858  R-----LKARISOSTKTFTPCER-----LEKRR-----TSPLFECTLRRSF 18932
           |||
           :|:
Db      1683  KENEKAKKSLEADLMQOEDPLAAERARRKQADLEKELEABLASLSGRNALQOBKRRLB 17422
           |||
           :|:
0Y      1893  RTGSVVRQKYEEO-----MIDPMWIK-----EEVSSARASTIDKRRKFOQMNQ---EQMA 1940
           |||
           :|:
           :|:
           :|:
Db      1743  ARIQLEEELEEEBOGNMEAMSDRRKAKTOAQEOLSNELATERSTAOKNESARQOLERONK 18020
           |||
           :|:
0Y      1941  KYMALIKEMWP-----YGSTLFDECK 1962
           |||
           :|:
Db      1803  ELRKSLHEBGAVKSKFRKSTJALAEAK 1829

```

Search completed: July 14, 2003, 18:27:25
Job time : 62.2242 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:11:27 ; Search time 6.99422 Seconds
(without alignments)
2620.804 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3252

Sequence: 1 MSTDQERQINTEYAVSLLEQ.....STGTEEPEDLGEVVALPPV 623

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfill1.pep:*

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	516.5	15.7	589	2	US-08-317-305-4
2	516.5	15.7	589	3	US-08-862-508-4
3	516.5	15.7	589	5	PCT-US95-12508-4
4	511.5	15.5	589	2	US-08-317-305-2
5	511.5	15.5	589	3	US-08-862-508-2
6	511.5	15.5	589	5	PCT-US95-12508-2
7	422.5	10.9	516	4	US-09-154-750A-86
8	357.5	10.9	642	4	US-09-337-307A-3
9	357.5	10.9	642	4	US-09-068-655-6
10	329	10.0	553	4	US-09-337-307A-4
11	309.5	9.4	530	1	US-08-307-499-29
12	309.5	9.4	530	4	US-09-299-268-29
13	234.5	7.1	500	1	US-08-307-499-20
14	234.5	7.1	500	4	US-09-299-268-20
15	195.5	5.9	478	4	US-09-137-223A-2
16	192.5	5.8	109	4	US-08-553-541B-7
17	192.5	5.8	109	4	US-09-268-202-7
18	167.5	5.1	803	4	US-09-063-035-2
19	135.5	4.1	706	1	US-08-074-967-2
20	135.5	4.1	706	2	US-08-553-541B-2
21	135.5	4.1	706	4	US-09-268-202-2
22	135.5	4.1	706	5	PCT-US94-06669-2
23	125.5	3.8	104	1	US-08-340-203A-6
24	125.5	3.8	104	2	US-08-452-427-6
25	125.5	3.8	104	3	US-09-085-407-6
26	125	3.8	106	1	US-08-340-203A-10
27	125	3.8	106	2	US-08-553-541B-5

28	125	3.8	106	2	US-08-452-427-10	Sequence 10, Appl
29	125	3.8	106	3	US-09-085-407-10	Sequence 10, Appl
30	125	3.8	106	4	US-09-268-202-5	Sequence 5, Appl
31	125	3.8	110	2	US-08-553-541B-9	Sequence 9, Appl
32	125	3.8	110	4	US-09-268-202-9	Sequence 9, Appl
33	124.5	3.8	569	4	US-09-137-223A-3	Sequence 3, Appl
34	123.5	3.8	108	2	US-08-553-541B-6	Sequence 6, Appl
35	123.5	3.8	108	4	US-09-268-202-6	Sequence 6, Appl
36	121.5	3.7	549	4	US-09-245-041-9	Sequence 9, Appl
37	121.5	3.7	1260	4	US-09-245-041-2	Sequence 2, Appl
38	121	3.7	104	1	US-08-340-203A-7	Sequence 7, Appl
39	121	3.7	104	2	US-08-452-427-7	Sequence 7, Appl
40	121	3.7	104	3	US-09-085-407-7	Sequence 7, Appl
41	120.5	3.7	1198	4	US-09-245-041-131	Sequence 131, App
42	120.5	3.7	1198	4	US-09-794-236-3	Sequence 3, Appl
43	120.5	3.7	1429	4	US-09-245-041-130	Sequence 130, App
44	120.5	3.7	2787	4	US-09-245-041-15	Sequence 15, Appl
45	115	3.5	105	1	US-08-340-203A-8	Sequence 8, Appl

ALIGNMENTS

```
RESULT 1
US-08-317-305-4
; Sequence 4, Application US/08317305
; Patent No. 5863744
; GENERAL INFORMATION:
; APPLICANT: Avraham, Shalom
; APPLICANT: Avraham, Hava
; APPLICANT: Groopman, Jerome E.
; TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING
; TITLE OF INVENTION: SAME
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
; STREET: Ten Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,305
; FILING DATE: 03-OCT-1994
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Ph.D., Kathleen A.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: DH-001XX
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-2290
; TELEFAX: (617) 451-0313
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-317-305-4

Query Match      15.7%  Score 516.5, DB 2, Length 589,
Best Local Similarity 25.5%  Pred. No. 1.3e-44;
Matches 155; Conservative 94; Mismatches 237; Indels 121; Gaps 17;

CY      11 TTYAASLEQLKFLFEQQLFTDIVLIVGTEBPCHKMTLATGSSYFRPMFMSGLSESXKT 70
DB      26 SSYADSVLTHLNLRLROQLFTDIVLILHAGNRTFPCRAVLACSRYPFAMFSGGLSESODS 85
```

```

0Y 71 VHLENN-VDAATLOIITAYATGNLMMNSTYQLETAECFLQVEVULQRCREYLLKKIN 129
Db 86 EVNPNDSIHEPEVLELDIYAISRVVINEENNESELEAGDMLEFQIDRCAAEFKENKH 145
0Y 130 AENCYRLLSFADLFSCEEELKOSAKRWEHKEFTAVYQDAFMQLSHLLDIDILSSDNLNVE 189
Db 146 PTNCGMLLSAHQOCTKUYELSMWRCLSNFOIRKNEDPDLQPDQMVVYQLSSSELEETE 205
0Y 190 KEETVREANMLLENTESRSQYLSVLSQIRIDALSEVTORAMFOGLPBNDSYVVQGL 249
Db 206 DERLYVESAMNNISYDLKKRYCYLPBELLQTVRLALPAIYLM-----ENVAMBEL 255
0Y 250 YKSMKPFPRRLGTMKE--EMMIFLASENPECSLYSVCYSPQ-----A 292
Db 256 ITKORK-----SEIYEBAIRCLKTLQNDGVNLSLCARPKTGHALFLLGGOTPMC 307
0Y 293 EKVY-----KLCSPPADL---HKVGVVTPDNDIYIAGQVYLNKNTKTNHSKTSKLOTA 343
Db 308 DKLYLVDOAKAKIIPRADI.PSPRKESACALGCKVYITG-----RSENGVSKD----- 357
0Y 344 FRTVNCFYW-FDAQONTWFPKTPMLFVRIKPSLVCCEGYIYALGDS-----VGGELN 395
Db 358 -----VWVYDTLHEBEMKAA.PMLVARFGHGAELKHCLYVVGHTAATGCLPASPSVS 410
0Y 366 RRTVERYPDKECBMTNVSP.LPCAMQMSAAV-----VYNDCLYMTNLNMYCY 442
Db 411 LKQVBOYDPTTKMTMVAP.LRBGVSSAAVYASAKLKLPAFGGTSVSHD-----KL.PYQOC 465
0Y 443 FPRBSWEMAMARQTSRSPASAAA.FGDKIFYIGLHATNSGIRLPSGTVDGSSVTEIY 502
Db 466 DOCEMRKMSPA.CPQPMRYTAAAVLGNQIFIM-----GDLEFSACAYKE 511
0Y 503 DVNKNEMQMAANI.PAKRYSDPCVRAVYVINSLSLVFMRETHLEBAKYTYQO-----YD 555
Db 512 NSETYQMTKVGVDYAKRMS---CHAVASGNKLYV-----GGVFGIQRCKTLDYD 559
0Y 556 LEIDRMS 562
Db 560 PTLDVNN 566

```

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 742-4214
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 589 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-862-508-4

```

Query Match	15.7%	Score 516.5	DB 3	Length 589
Best Local Similarity	25.5%	Pred. No. 1.3e-44		
Match 155	94	Mismatches 237	Indels 121	Gaps 17

```

QY 11 TEXVATLEBQLLFTFEQOLFTDVLIVYBGEFEPCHKNVLAITCSSYFPAFMSGLSSSKOT 70
Db 26 S$TADSVLTHLNLKQORLFTDVLHAGNRTFPGHRAVLACSSYFPAFMSGGGLSSKODS 85
QY 71 HVHARS-VDAATLOIITVAYATGNLAMDSTVEOLYETACGLQVEDVLQCREYLLIKIN 129
Db 86 EVAFND$SIHPEVLELIDLOVAYS$RVIINEBNA$SLLEAGDMLERQODIRDA$CAFLEKNLH 145
QY 130 AENCVRFL$PADLFCSEBILKOSAKMVEHKTAVYHODAFMQLSHDLIDISSDNLNVE 189
Db 146 PTMCLMGLL$SAHOCTKLYEL$MCMCLSNFOTIRKHEDFLOLPODMVVOILL$SELETE 205
QY 190 KEETVBEA$MLLEYN$TESRQYLS$VLSOIRIDLS$VTQRAWFQGLP$END$VVVOGL 249
Db 206 DE$LVYESAMNISTDLKRYCYLPELLOQYRLMLLPAIYLM-----ENVAMEBL 255
QY 250 YK$MPF$K$RBLG$MKE--EMMIFL$ASE$N$C$LSY$SVCSY$PQ-----A 292
Db 256 ITKORK-----SKEIYE$IRCKLILONDGVV$TSLCARPKTHALFLLGQTFMC 307
QY 293 EKY-----KL$SP$ADL-----HKGYGVTP$DNDIYIAGQVPLKXNKT$H$K$T$K$QTA 343
Db 308 DKLYLV$DQAKKIIIPADIP$E$RKEFS$CAIGCKVYITG-----R$GENGV$XD----- 357
QY 344 FRTVNCFYW-FDAOONTW$PKT$PMLFV$IKR$SLV$C$EGYIYAIGDS-----V$G$ELN 395
Db 358 -----VWVYDTLH$E$MSKAP$MVLVAFH$G$AL$H$KCLYVVG$HTAATGCLPASP$VS 410
QY 396 RRTVEY$YDTEK$E$B$T$M$V$B$PLC$A$M$O$M$A$V-----VH$D$CIY$M$T$L$N$M$Y 442
Db 411 LKQVEYEDPT$TKM$T$M$V$A$PL$E$G$V$S$N$A$V$S$AK$L$P$A$F$G$T$V$S$H$D-----KL$P$K$O$C 465
QY 443 FPRSD$W$VEMAR$Q$T$R$S$P$A$A$A$B$G$D$K$IF$YIG$GL$H$AT$N$G$IR$P$E$G$T$V$D$G$S$V$T$E$Y 502
Db 466 DQ$EN$M$S$P$A$C$P$O$P$M$R$Y$T$A$A$V$I$G$N$O$IF$M$G-----G$D$E$F$S$A$C$A$Y$K 511
QY 503 DVN$K$E$M$K$A$N$A$IP$A$K$Y$S$D$P$C$V$R$A$V$V$S$N$L$C$V$F$M$E$T$H$L$N$E$A$K$Y$T$V$Q-----YD 555
Db 512 N$E$T$Y$O$M$T$K$V$G$V$T$A$K$R$M$S-----CH$A$V$A$G$N$K$L$Y$V-----G$G$Y$F$I$Q$R$C$K$T$L$D$C$YD 559
QY 556 LEIDR$S 562
Db 560 PTL$D$V$N 566

```

RESULT 3
PCT-US95-12508-4
Sequence 4, Application PC/TUS9512508
GENERAL INFORMATION:
APPLICANT: Avraham, Shalom
APPLICANT: Avraham, Hava
APPLICANT: Grooman, Jerome E.
TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING SAME
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lahive & Cockfield
STREET: 60 State Street, Suite 510
CITY: Boston

STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/12508
FILING DATE: 29-SEP-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/317,305
FILING DATE: 03-OCT-1994
ATTORNEY/AGENT INFORMATION:
NAME: Silverl, Jean M.
REGISTRATION NUMBER: 39,030
REFERENCE/DOCKET NUMBER: NER-259PC
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 589 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-12508-4

Query Match 15.7%; Score 516.5; DB 5; Length 589;
Best Local Similarity 25.5%; Pred. No. 1.3e-44;
Matches 155; Conservative 94; Mismatches 237; Indels 121; Gaps 17;

11 TEVAVSLLEQLKFLFEOQLFTDVLIVEGTEPFCNKVTLATCSSYFRAMFMSGLSSKOT 70
26 SSYADSVLTHLNLRLQORLFTDVLHAGNRTPCHRAVLAACSRFEAMFSGGLKESQDS 85
71 HVLHLSN-VDAATLQIITFYATGNLAMDSTVEOLYETACFLQVEDVLCRCREYLIIKIN 129
86 EVNPNDSIHPEVLELLDLYAISRVIINENNESLLEAGDMLEFQDIRACAEFLKINH 145
130 AENCVRLLSPADLFSCCEELKQSAKRVNHEKFTAVVYHQAFLMQLSHDLLIDILSSDNLYE 189
146 PTNCLGMLLSDAHQCTLYELSWRMCLSNFQIRKNEDELQLPQDMVYQLLSSELEFTE 205
190 KEETREAAAMLLENTSRSQYLSVSQIRIDALSIYTORAMFQGLPNDKSVVYQGL 249
206 DERLYVESAMNWSIDLKRKYCYLPQLQTVRLALLPAILYM-----ENVAMEEL 255
250 YKSMKPFKPRLGMTKE--EMMIFIASSENPCSLYSVCYSPQ-----A 292
256 ITRKQR-----SKEIYBEAIRCKLKLQNDGVVTSLCARPKRTGHALPILGGOTFNC 307
293 EKVY-----KLCSPADL--HKVGTVPDNDIYIAGQVPLKNTKTNHSTKSLQTA 343
308 DKLYLVDOQAKKEIIPKADIPSPKKEFSACALGCKVYITG---RSENGVSKD----- 357
344 FRTVNGFYV-PDAQQNTWFPKTPMLFVRIRKPSLVCEZYIYAGGS-----VGEELN 395
358 -----VVVYDYLHEEWSKAAPMLVARFGHSGAELKCLYVGGHTAATGCLPASPVS 410
396 RRTVERYTEKDEMTWVSPLPACAMQWSAAV-----VVHDCIYVMTLNMICY 442
411 LKQVEYDPTTKMTKTVAPLREGVSNAAYVSATLKLFAGGTSVSHD-----KLPRVQCY 465
443 PRSDSVEMAMKQTSRSFASAAAFQDKI FYIGGLHIAITNSGIRLPSTGTVDSGVTEIY 502
466 DDCENRWSVPATCPQPMRYTAAVLAALNGQIFIMG-----GDTFSACSAKYKF 511
503 DVNKNQKMAAMNIPARKYGDPCRAVVISNLSVPRKEETHLMBRAKYVYQ-----YD 555
512 NSEYIOWTVYGVDTAKRMS--CHAVASGNKLYV-----GGYFGIQRCKTILDYD 559

QY 556 IELDRLWS 562
DB 560 PTLDVWN 566

RESULT 4
US-08-317-305-2
Sequence 2, Application US/08317305
Patent No. 5863744
GENERAL INFORMATION:
APPLICANT: Avraham, Shalom
APPLICANT: Avraham, Hava
APPLICANT: Groopman, Jerome E.
TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING
TITLE OF INVENTION: SAME
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Weingarten, Schurgin, Gagnebin & Hayes
STREET: Ten Post Office Square
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/317,305
FILING DATE: 03-OCT-1994
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Williams Ph.D., Kathleen A.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: DH-001XX
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 542-2290
TELEFAX: (617) 451-0313
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 589 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-317-305-2

Query Match 15.5%; Score 511.5; DB 2; Length 589;
Best Local Similarity 25.4%; Pred. No. 4.3e-44;
Matches 154; Conservative 94; Mismatches 238; Indels 121; Gaps 17;

11 TEVAVSLLEQLKFLFEOQLFTDVLIVEGTEPFCNKVTLATCSSYFRAMFMSGLSSKOT 70
26 SSYADSVLTHLNLRLQORLFTDVLHAGNRTPCHRAVLAACSRFEAMFSGGLKESQDS 85
71 HVLHLSN-VDAATLQIITFYATGNLAMDSTVEOLYETACFLQVEDVLCRCREYLIIKIN 129
86 EVNPNDSIHPEVLELLDLYAISRVIINENNESLLEAGDMLEFQDIRACAEFLKINH 145
130 AENCVRLLSPADLFSCCEELKQSAKRVNHEKFTAVVYHQAFLMQLSHDLLIDILSSDNLYE 189
146 PTNCLGMLLSDAHQCTLYELSWRMCLSNFQIRKNEDELQLPQDMVYQLLSSELEFTE 205
190 KEETREAAAMLLENTSRSQYLSVSQIRIDALSIYTORAMFQGLPNDKSVVYQGL 249
206 DERLYVESAMNWSIDLKRKYCYLPQLQTVRLALLPAILYM-----ENVAMEEL 255
250 YKSMKPFKPRLGMTKE--EMMIFIASSENPCSLYSVCYSPQ-----A 292
256 ITRKQR-----SKEIYBEAIRCKLKLQNDGVVTSLCARPKRTGHALPILGGOTFNC 307
293 EKVY-----KLCSPADL--HKVGTVPDNDIYIAGQVPLKNTKTNHSTKSLQTA 343

Db 308 DKLVLVDQAKKEIIIPKADIPSPKESFACAIQCKYITGG-----RSGENGVSKD----- 357

QY 344 FRTVNCFW-FDAQONTWPKTPEMLFVRIKPSLVCCGYIYAGDS-----VGGELN 395

Db 358 -----VWVYDTLHEMSKAAPMLVAFHGSAELKHCILYVGGHTAATGCLPASPSYS 410

QY 396 RRTVERYTEKDEMTWVSPPCAMQMSAAV-----VHDCIYVMTLNLMYCY 442

Db 411 LKQVEHYDPTINKMTWVAPLREGVSNAAVSAKLKLPAGGTSVSHD-----KLPRKYQCY 465

QY 443 PRSDSVEMANQTSRSPASAAAFGDKFYIGGLHATNSGIRLPSGTVDSSVTEIY 502

Db 466 DQENRMTVPATCPQPMRYTAAVIGNQIFIMG-----GDTEFSACSAAYKF 511

QY 503 DVNKNEMKMAANIIPAKRYSDPCVRAVVISNSLCVFMRETHLNERAKYTYO-----YD 555

Db 512 NSEYQMTKVGDTYAKRMS-----CHAVASGNKLYV-----GGYFGIQCKTLDCYD 559

QY 556 LELDRWS 562

Db 560 PTLDVWN 566

RESULT 5

US-08-862-508-2

Sequence 2, Application US/08862508

Patent No. 6066451

GENERAL INFORMATION:

APPLICANT: Avraham, Shalom

APPLICANT: Avraham, Hava

TITLE OF INVENTION: NOVEL NEURAL CELL PROTEIN MARKER RR/B AND DNA

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lahive & Cockfield, LLP

STREET: 28 State Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/862,508

FILING DATE: 23-MAY-1997

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/317,305

FILING DATE: 03-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: Silveri, Jean M.

REGISTRATION NUMBER: 39,030

REFERENCE/DOCKET NUMBER: NER-259DV

TELECOMMUNICATION INFORMATION:

TELEPHONE: (617) 227-7400

TELEFAX: (617) 742-4214

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 589 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-862-508-2

Query Match 15.5%; Score 511.5; DB 3; Length 589;

Best Local Similarity 25.4%; Pired. No. 4.3e-44;

Matches 154; Conservative 94; Mismatches 238; Indels 121; Gaps 17;

QY 11 TEAVSVLLBQKLFYEQQLFTDVLVVEGTFEFCMKVTLATGSSYRRAMMSGISLSSKOT 70

Db 26 SSYADSVLTLLNLIRQORLEFTDVLHAGNTPFCBRAVLAACSRFEAFMSGGLKESODS 85

QY 71 HVHLRN-VDAATLIQIITTYAATGNLAMDSTVEQYETACFQVEDVLQRCREYLIIKIN 129

Db 86 EYVFNDSIHPEVLELLDQVAYSRIYINENNESLLEBGMLEFPQDINDACAEFLKINH 145

QY 130 AENCVRLLSFADLFCEBEIKQSAKMEVHEKFTAVYHQDAFMQLSHDLIDLISSONLVE 189

Db 146 PTNCLGMLLSDAHOCTKLYELSMWCLSNFQTIKNEFDLQLPQDMVQQLSSSELETE 205

QY 190 KEETVEAAMLLENTERSOYLSYSQIRIDLSBVTQAMQFQJLPNDKSVYVQGL 249

Db 206 DERLYESALINMISTDLKRCYCLPELQTVRLALPLIYLM-----ENVAMEBL 255

QY 250 YKSMPEFPRRLGWTKE--EMMIFTEASENPPCSLYSSVCYSPQ-----A 292

Db 256 ITRKQK-----SKEIYEAI RCKLKLQNDGVVTSICARPRTHGALFLGGOTFWC 307

QY 293 ERYVY-----KLCSPADI--HXYGVTVTPNDIYIAGQVPLKNTKTNSKTSKLQTA 343

Db 308 DKLVLVDQAKKEIIIPKADIPSPKESFACAIQCKYITGG-----RSGENGVSKD----- 357

QY 344 FRTVNCFW-FDAQONTWPKTPEMLFVRIKPSLVCCGYIYAGDS-----VGGELN 395

Db 358 -----VWVYDTLHEMSKAAPMLVAFHGSAELKHCILYVGGHTAATGCLPASPSYS 410

QY 396 RRTVERYTEKDEMTWVSPPCAMQMSAAV-----VHDCIYVMTLNLMYCY 442

Db 411 LKQVEHYDPTINKMTWVAPLREGVSNAAVSAKLKLPAGGTSVSHD-----KLPRKYQCY 465

QY 443 PRSDSVEMANQTSRSPASAAAFGDKFYIGGLHATNSGIRLPSGTVDSSVTEIY 502

Db 466 DQENRMTVPATCPQPMRYTAAVIGNQIFIMG-----GDTEFSACSAAYKF 511

QY 503 DVNKNEMKMAANIIPAKRYSDPCVRAVVISNSLCVFMRETHLNERAKYTYO-----YD 555

Db 512 NSEYQMTKVGDTYAKRMS-----CHAVASGNKLYV-----GGYFGIQCKTLDCYD 559

QY 556 LELDRWS 562

Db 560 PTLDVWN 566

RESULT 6

PCT-US95-12508-2

Sequence 2, Application PC/TUS9512508

GENERAL INFORMATION:

APPLICANT: Avraham, Shalom

APPLICANT: Avraham, Hava

TITLE OF INVENTION: NOVEL PROTEIN MARKER AND DNA ENCODING SAME

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Lahive & Cockfield

STREET: 60 State Street, Suite 510

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/12508

FILING DATE: 29-SEP-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/317,305

FILING DATE: 03-OCT-1994

ATTORNEY/AGENT INFORMATION:

NAME: Silveri, Jean M.

REGISTRATION NUMBER: 39,030

REFERENCE/DOCKET NUMBER: NER-259PC

TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 589 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-12508-2

Query Match 15.5%; Score 511.5; DB 5; Length 589;
Best Local Similarity 25.4%; Pred. No. 4.3e-44;
Matches 154; Conservative 94; Mismatches 238; Indels 121; Gaps 17;

11 TEVAVLLLEQLKLFYEQQLFTDVLVLEGTEPCHKMTATGSSYFRAMFMSGLSSKOT 70
26 SSTADSVLTHNLRLRQRLFTDVLHAGNRTFPCRAVLACSRIFEAMFSGGLSKESQDS 85
71 HVALRN-VDAATLQIITVAYTGNLAMDSTVEQLYETACFLOVEDVLQRCREYLKIKIN 129
86 EVNFDNSIHPEVLELLDVAYSRVIIENEMASLEADIMLEFODIRDACAFLEKNLH 145
130 AERCVRLSFADLFSCBELKQSAKRVNFKFTAVYHODAFMQLSHDLIDILSSDNLNVE 189
146 PTNCLMLLSDAHQCTKLYELSMWCLSNFQIRKNEPFLQPODMVQLLSSELETE 205
190 KEETVEAAMLMEYTESRSQYSSVLSQIRIDALSEVTOGRAMFGQLRPNDKSVVVOGL 249
206 DELVYESAINWSTYDKKRYCYLPBELLQTVRALLPALYLM-----ENVAEEL 255
250 YKSMPEKPRKMTKE--EMMIFIEASSENPCSLYSVCYSPQ-----A 292
256 ITRKOR-----SKEIYBEAIRCKLKIQNDGVVTSICARPKTGHALFLGQTPMC 307
293 EKYI-----KLCSPADL---HKVGVTPDNDIYIAGGVPLKNTKTNHSTSKLQTA 343
308 DKLYLDOKAKELIPRADIPSPRKEFSACAIQCKVYITG---RSENGVSD----- 357
344 FRVNCFFV-FDAQONTWPKTPTMLFVRIKPSLVCEGYIATIGDS-----VGGELN 395
358 -----VVVYDILHEEMSGAAMPVLARFGHSAELKELCYVGVGHTAATGCLPASPSVS 410
396 RRTVERYTEKDEMTWVSPRPCAMQWSAAV-----VHDCIYVTLNLMYCY 442
411 LQVEHYDPTINKMTVAIPREGVSNAVSAKLKLPARGCTSVSD---KLPRVQCY 465
443 PPRSDSVEMANRQTSRSPASAAAFDDKIFYIGLHIATNSGIRLPSGTVDSSVTEIY 502
466 DDCENRWTPATCPQPMRYTAAAVLGNQIFIMG-----GDTEFSACSAVXF 511
503 DVNKNEMKMAANI PARYSDPCRAVVISLGVMEETHLNBRKAYVYQ-----VD 555
512 NSETTYMTKGVDTAKRMS---CHAVASGNKLYV-----GGYFGIQRCKTLDCYD 559
556 LELDRMS 562
560 PTLDVVN 566

RESULT 7
US-09-154-750A-86
Sequence 86, Application US/09154750A
Patent No. 6432640
GENERAL INFORMATION:
APPLICANT: Vogelstein, Bert
APPLICANT: Kinzler, Kenneth
APPLICANT: Polyak, Kornelia
TITLE OF INVENTION: p53-Induced Apoptosis
FILE REFERENCE: 1107,75357
CURRENT APPLICATION NUMBER: US/09/154,750A
PRIOR FILING DATE: 1998-09-17
PRIOR APPLICATION NUMBER: 60/059,153

PRIOR FILING DATE: 1997-09-17
PRIOR APPLICATION NUMBER: 60/079817
PRIOR FILING DATE: 1998-03-30
NUMBER OF SEQ ID NOS: 93
SOFTWARE: FastSeq for windows Version 3.0
SEQ ID NO 86
LENGTH: 516
TYPE: PRT
ORGANISM: Homo sapiens
US-09-154-750A-86

Query Match 12.8%; Score 422.5; DB 4; Length 516;
Best Local Similarity 24.1%; Pred. No. 6.1e-35;
Matches 151; Conservative 87; Mismatches 220; Indels 169; Gaps 21;

11 TEVAVSLLEQLKLFYEQQLFTDVLVLEGTEPCHKMTATGSSYFRAMFMSGLSSKOT 70
26 SSTADSVLTHNLRLRQRLFTDVLHAGNRTFPCRAVLACSRIFEAMFSGGLSKESQDS 85
71 HVALRN-VDAATLQIITVAYTGNLAMDSTVEQLYETAC-----PLOVEDVLQ 119
86 EVNFDNSIHPEVLELLDVAYSRV-----IHQL-EGKCRSLGSLVTCSPFDIRRA 138
120 CREYLKIKINAENCVRLSFADLFSCBELKQSAKRVNFKFTAVYHODAFMQLSHDLID 179
139 CAEFLKKNLHPNCLGMLLSDAHQCTKLYELSMWCLSNFQIRKNEPFLQPODMVVO 198
180 ILSSDNLNVEKETVEAAMLMEYTESRSQYSSVLSQIRIDALSEVTOGRAMFGQLRP 239
199 LLSSELETERLVEBSAINWSTYDKKRYCYLPBELLQTVRALLPALYLM----- 250
240 NDKSVVVOGLYKSMPEKPRKMTKE--EMMIFIEASSENPCSLYSVCYSPQAEKYK 297
251 --ENVAEELITRKOR-----SKEIYBEAIRCKLKIQNDGVVTSICARP----- 294
298 LQSPPADLHKVGVTPDNDIYIAGGVPLKNTKTNHSTSKLQTAFTVNC--PYWEDA 355
295 -----KTS-----HALFLGQ-----TFMCDKLYLVQ 318
356 QONTWPKTPTMLFVRIKPSLVN---CEGYIATIGDSVGGELNRTVER---YDTRKE 408
319 KKEKELPKADISPR-KEFSACAIQCKYI-----TGRSENGVSDVVVYDILHEE 370
409 WTVSPRLPCAMQWSAAVVHDCIYVTLNLMYCYFPRSDSVEMANRQTSRSPASAAAG 468
371 WSKAAPMLVARKGHSAELKCLYV----- 395
469 DKIFYIGLHIATNSGIRLPSGTVDSSVTEIYDVNKNEMKMAANIPAKRYSDPCRAV 528
396 -----VGHHTAATGCLPASPSVSLK---QVEHYDPTINKMTVAIPRRRYN--CAQV 444
529 VLSNGLCVMEETHLNBRKAYVYTOYDELDWLSRHLSEKVLNDLGRDFCTYCKLYP 588
445 SAKLKLPAFGGTGSVSHDKLPKVOQ--YDCCENRW-----TVPATCP 483
589 SCLBSPMKPPTYLPSTDGTEBEFLDG 615
484 -----QPMRIHQASCPSGTQDFLLMG 505

RESULT 8
US-09-337-307A-3
Sequence 3, Application US/09337307A
Patent No. 6432692
GENERAL INFORMATION:
APPLICANT: Bradfield, Christopher A.
APPLICANT: Carver, Lucy A.
APPLICANT: Dunham, Elizabeth E.
TITLE OF INVENTION: Sensitive Biosay For Detecting Agnists of The Aryl Hydrocarbon
FILE REFERENCE: MAR0105
CURRENT APPLICATION NUMBER: US/09/337,307A
PRIOR FILING DATE: 1999-06-21
NUMBER OF SEQ ID NOS: 19


```

? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 3
? LENGTH: 642
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURES
? NAME/KEY: misc feature
? OTHER INFORMATION: /note = "AraJ"
US-09-337-307A-3
```

Query Match	10.9%;	Score 357.5;	DB 4;	Length 642;
Best Local Similarity	18.8%;	Pred. No. 5e-28;		
Matches 134;	Conservative 114;	Mismatches 270;	Indels 195;	Gaps 22;

```

0Y 4 ODERINTENVAUSLLEIOLFLFYEQOIFETDIIVIVESTEPFCHMTATCSTYRPAFMG 63
Db 10 EDENITESSVA-----KUNALKRSGQFCVRIQVCEHMLAHRAVLACCSPLYLFFNSD 64
0Y 64 LSESQTHVHLNNDATLQIITITATYTNLAMDSTVEOLYETACFLOVEDVLOCREY 123
Db 65 SDPHGISHHKFPDDLDEAVEVELVNTAVYTQLKADKELVXDVYSAAKLRKDRVKQVCGY 124
0Y 124 LIKKIINAEVCVALLSPADLFSCSEBLQSAKRWENHFTAVUHODAFMOLSHDILILSS 183
Db 125 LLSRBDVYSCICYRRFPASCMDSRLLINKDAVYIQEHLLOISEEEFLKLPRLKLEVMLE 183
0Y 184 DNLAVEKEETVEAAMLWLEYNTESTRQYILSSVLOIR-----IDALSEY 228
Db 184 DNVCLPSNGKLTXYKVLVNWQSRIMWEGDSELEIMEEVQLYYSADHKLLDGNLLDGOAEV 243
0Y 229 -----TORAMFGCLPNDKS----- 243
Db 244 FGSDDDHIOFVOKKPRENGHGHQIOLSSSTGCJSSPNATVQSPRHEMKIYASEKTSNNTLY 303
0Y 244 ---VVVOGLY-----XSMPEFFK-----PRLGMTK 265
Db 304 CLAVLDGIFCVILFHGRNSPOSSPTPIKLSLSPFMOQDELIEKPMSPMQYARSGGLG 363
0Y 266 EEM---MFIEMSENPCGLSSVCSPOAEKYKLCSP---PADLHXGTVTTPNDIY 319
Db 364 AEMNGKLLIAGGYNNEEC-LRTVECNPHPTDH-WSLAPMRTRERABFQMAVLN---GOLX 418
0Y 320 IAGGOVPLKNTKNTKNSKTSKQTAFTFVVCFCWFYFDQOQWTFPEKTEPMLFVIRIPSLVCS 379
Db 419 VVGG-----SNGHSD-----DLSCGEYNDINIDMIPVPELRTNRCAAGCALN 462
0Y 380 GYIYVAGGDSVGGELNRRTRVERDYDEKEDENTVSPJPCAMOWSAVAVVDHCIVM----- 434
Db 463 GKLTYVGGDDPYGOKGLKNKCDVDFPPTKMLTSCAPINIRRHOSAVELGGLYLIIGABE 522
0Y 435 --TLNLMCYFPRSDSWMAMARQOTRSFASAAAFEDKIFYJGGLHIAATSGIRLPSGY 492
Db 523 WNCILMTVERIYENBNTWTLLAPMNVARRGCAVALNGKLFVCCGF----- 567
0Y 493 DGSSV--TVEIYDVNKNEMKMANIPAKYSDPCRAVVIYNSLCVFMRETHINERAKYV 550
Db 568 DGSHAISCVEMHDPTRNEWMKMGNMNTPSPSN--AGIATVGTNI-----YA 610
0Y 551 TYQVLELDRWLSLRQHISERVLMDGRDRCTYGGKLYPSCLSESPMKPPTLYF 603
Db 611 VGGFD-----GNEFLNTV-EVYN--LESSEWSPYTKIF 640

```

RESULT 9
US-09-337-307A-4
; Sequence 4, Application US/09337307A

Patent No. 6432692
GENERAL INFORMATION:
APPLICANT: Bradfield, Christopher A.
APPLICANT: Carver, Lucy A.
APPLICANT: Dunham, Elizabeth E.
TITLE OF INVENTION: Sensitive Biossay For
FILE REFERENCE: WARF0105

Detecting Agonists Of The Aryl Hydrocarbon

```

:
:
: CURRENT APPLICATION NUMBER: US/09/337,307A
:
: CURRENT FILING DATE: 1999-06-21
:
: NUMBER OF SEQ. ID NOS: 19
:
: SOFTWARE: PatentIn version 3.1
:
: SEQ. ID NO. 4
:
: LENGTH: 642
:
: TYPE: PRT
:
: ORGANISM: Mus musculus
:
: FEATURE:
:
: NAME/KEY: misc feature
:
: OTHER INFORMATION: /note = "ARA3"
:
US-09-337-307A-4

```

Query Match	10.9%;	Score 357.5;	DB 4;	Length 642;
Best Local Similarity	18.8%;	Pred. No. 5e-28;		
Matches 134; Conservative	114;	Mismatches	270;	Gaps 195;
				Indels 22

QY	4	QDRQJQNTBNAVSLEEQULFYEQOOLFNDIVLIVEGNEFFCHKNVLLTCCSSYFRAMFMSG	63
Dd	10	EDENFLEISSA-----KLNALKKSQOFCDVRLQVCGHMLHRAVLACCSPYLEIFNSD	64
QY	64	LSSKQTHVHLKNVDAITQIITATYAGNLAMDSTYEQULYRACFLQVEDULQRCREY	123
Dd	65	SDPHGISHYKFDLDEAVEVLNAYALOLKADHELKDVYSAKKLKMDRVQVQGDY	124
QY	124	LIRKIANEANCVRLLSPADLFSCSEELKQSAKRVEHKEFPAYVNOAFQMLSHDILLIDLSG	183
Dd	125	LISRMVTSCTISYRNASCSMGDSRLLKNVDAYIQHLLQISEEBEFLKLR-LKLEVMLE	183
QY	184	DNINVEKETREAAMLWLEYNTESRSQYLSVLSQIR-----IDALSEV	228
Dd	184	DNYCLPSNGKLYKTVINWQORSIWENGDSELBELMEVQTLYSADHKLLDGNLDGOAEV	244
QY	229	-----IQRAMFOGLPBNDS-----	243
Dd	244	FGSDDDHIOFQVKKPPRENGHKQIOSSSTGCCLSFNATVQSPKJENKIVASEKTSNNTYL	303
QY	244	--VVOGLY-----KSMKPFK-----BRLGMYK	265
Dd	304	CLAVLDGIFCVLFIHGRNQPSSPTSTPLKSLKSFEMQODELIEKPMSPQVARSIGLT	363
QY	266	EEW---MIFLEASSENPCSLSSVCTYSQAENVYLCSP---PADLHKVGVVTPDDIY	319
Dd	364	AENNGKILAAAGYNREEC-LRTVECYNPTDH-WFLAPMRTPRARQOMAVLM--GOLY	418
QY	320	IAGGVPALKATKTNHSKTSKLOTAERTVNACFWPAQONTFPRKTPMLFYAIRKPSLVCE	379
Dd	419	VVGSG-----SNGSHD-----DLSCEGMDYSNIDIMVPIBELRTNRCNAGVCLN	466
QY	380	GIIYAIAGDSVGGELNRTYERYDIEKCBWTVSFLPCAMQMSAAVNVHDCIYV----	433
Dd	463	GKLIYVGGSDPYQOKLKNKCDVFDVTKLMTSCAPLIRHQSAVCELGAGLYLIGGAES	522
QY	435	--TLNIMVCFPPRSDGVMEMARQTSRSPASAAAGDKIFYIGLHILATNSGIRLPSGTV	492
Dd	523	WNCNLTVERYNPENNNTWTLIAPMNVAARKGAYVANGLFLVCGGF-----	566
QY	493	DGSSV--TWEIYDVNKEKMAANIPAKRYSDPCRAVAVISNSLCEVRETHLNERAKYV	556
Dd	568	DGSHAISCEVEMHDPTEJENWKMGNTSPRSN---AGIATVGNTI-----YA	610
QY	551	TYQYDLELRMSLRQHISERVLMDGDPDCTGVKLYSCLESSEPKPPTLYF	603
Dd	611	VGGD-----GNEFLNTV-EVYN--LESNBMSPYTIKF	640

RESULT 10
US-09-068-655-6
; Sequence 6, Application US/09068655A

```

; GENERAL INFORMATION:
; APPLICANT: JACKSON, Ronald James
; TITLE OF INVENTION: METHOD OF PRODUCING ALPHA 2,3-SIALYLTRANSFERASE

```

```

1 FILE REFERENCE: 026579-186
2 CURRENT APPLICATION NUMBER: US/09/068,655A
3 CURRENT FILING DATE: 1998-08-31
4 EARLIER APPLICATION NUMBER: AU PN 6587
5 EARLIER FILING DATE: 1995-11-15
6 EARLIER APPLICATION NUMBER: PCT/AU96/00725
7 EARLIER FILING DATE: 1996-11-15
8 NUMBER OF SEQ ID NOS: 13
9 SOFTWARE: PatentIn Ver. 2.0
10 SEQ ID NO 6
11 LENGTH: 553
12 TYPE: DRT
13 ORGANISM: Myxoma virus
14 US-09-068-655-6

```

Query Match	10.0%;	Score 329;	DB 4;	Length 553;
Best Local Similarity	21.2%;	Pred. No. 3.6e-25;		
Matches 131;	Conservative 85;	Mismatches 231;	Indels 172;	Gaps 21;

```

QY      13 YASLLLEQKLFEEQOLFDFDIYLIY-EGNEPFCMGVLTATSSYFRRAMFSGISEKQTH 71
Db      5 YAIRLLESTIRNLODKTKTCLCDVTLVYDDVDSIAHKLIIASSSTYFEYWFHSDFEKKRVN 64

QY      72 VHLRNDAATLQIIITYAYTGNLMMNDSTFVEQYETACFLQVEDYLQCREYLIKKINAE 131
Db      65 INV-CVEYRALHLINFIYSGLTRLTDPTDVCITLVAADYILQILEASELANFIILARIAE 123

QY      132 NCYRLLSPADLSCEELKOSAKRWBEKFTAYTHODAFMQLSHDLLIDILSSDNLNVEKE 191
Db      124 NCLHAYEFKSKRYNRHRIFFNVVITTTIIHNVSATLRQPNFKYTELCDLQNIILSSDPLANIYDE 183

QY      192 ETYREAAMLLEVENTSRSQVILSYLSQIRIDASE-----VTGRAMQGL--- 237
Db      184 DVCAVVLVTMLKON--NMEDCPSVLLEQVRMSLSMSYKULLLTPCTIRNKRIVQSLAKL 241

QY      238 -----PNDKSV----- 244
Db      242 DHSRRPPTQGLISTIGRKRYDDITSSPEVELSPYDDWCTVSYLPTHRQFPFYAVUDFV 301

QY      245 --VVGGL-----YKSNPKFKFRLG--MTKEEMMIFLEASSENPCG 281
Db      302 VVYVVGGLQDSVSVASVSVDYKTNMEKECPPLKSPRHGCGLVLRDKLIVIGKGRN--S 359

QY      282 LVSSVCY-SPOEKRYKLCSPPADLHKGTGVTTPNDIYYAGGQVPLKNTKTNHSTKSKL 340
Db      360 YLKDQVYMPPTAYATWKLCSLREARTNNGAAVY--RNKYTYTIGG---IRSVY---EBSRL 411

QY      341 QTAFTFVNCFYMFDDAQNTWFPK----TPMLFVRIRKESLVNCCGGIYALIGDVSGBELNR 396
Db      412 E-CIDTVECL-----QNNKTVAKKSLPEPKACLAAP---YKRFITYAAGGYALNG---R 458

QY      397 RTVERDYTEKEDWTWSPRPCAMQMSAAVVHDCIYWTLLNLMTCYFRSPDSWENAMRQ 456
Db      459 GTV-----YKNTLTLMNVELDWDFYLPME 485

QY      457 TSSRFSASAAFGDKIFYIGGLHAIATNSGIRLPSGTGVDSSVTEIYDVNKNEMKMANIP 516
Db      486 LSRNDASLCLVGLHDLIVYVGGF-----VGSGYTNSVERKYNHKNKTNMERIIPCK 532

QY      517 AKRYSDPCRAVAVISSLG 535
Db      533 SPKYG---HGSVVLNHEC 547

RESULT 11
US-08-307-499-29
Sequence 29, Application US/08307499
Patent No. 5651972
GENERAL INFORMATION:
APPLICANT: Moyet, Richard W.
APPLICANT: V4 uela, Eladio
APPLICANT: GIBDB, E.P.U.
TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a

```

```

1  TITLE OF INVENTION: Live Vaccine Vector
2  NUMBER OF SEQUENCES: 60
3  CORRESPONDENCE ADDRESS:
4  ADDRESSEE: David R. Saliwanchik
5  STREET: 2421 N.W. 41st Street, Suite A-1
6  CITY: Gainesville
7  STATE: Florida
8  COUNTRY: U.S.A.
9  ZIP: 32606
10 COMPUTER READABLE FORM:
11 MEDIUM TYPE: Floppy disk
12 COMPUTER: IBM PC compatible
13 OPERATING SYSTEM: PC-DOS/MS-DOS
14 SOFTWARE: Patent in Release #1.0, Version #1.25
15 CURRENT APPLICATION DATA:
16 APPLICATION NUMBER: US/08/307,499
17 FILING DATE:
18 CLASSIFICATION: 435
19 PRIOR APPLICATION DATA:
20 APPLICATION NUMBER: US 07/908,241
21 FILING DATE: 1-JUL-1992
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: US 07/908,630
24 FILING DATE: 29-JUN-1992
25 CLASSIFICATION: 435
26 PRIOR APPLICATION DATA:
27 APPLICATION NUMBER: US 07/342,212
28 FILING DATE: 21-APR-1992
29 CLASSIFICATION: 435
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Saliwanchik, David R.
32 REGISTRATION NUMBER: 31,794
33 REFERENCE/DOCKET NUMBER: DF35.1.FWCCI
34 TELECOMMUNICATION INFORMATION:
35 TELEPHONE: 904-375-8100
36 TELEFAX: 904-372-5800
37 INFORMATION FOR SEQ ID NO: 29:
38 SEQUENCE CHARACTERISTICS:
39 LENGTH: 530 amino acids
40 TYPE: amino acid
41 TOPOLOGY: linear
42 MOLECULE TYPE: protein
43 US-08-307-499-29
44
45 Query Match 9.4%; Score 309.5; DB 1; Length 530;
46 Best Local Similarity 21.7%; Pred. No. 3,5e-23;
47 Matches 125; Conservative 98; Mismatches 215; Indels 139; Gaps 22
48
49 QY 16 SLLEQLKLFYEQQLFTD-IVLYIEGTEFPCHKVNLATCSSYFRAMFSGLSSESQTHVH 73
50 Db 3 SLRLRLHDFFGHGMCDIKIYISINNKTISAHRLILSMYSKIFYNIFNSDFDKNDDELY 62
51 QY 74 LRVNDAATLQIIIVAYATGNLANMDSYEQLYETACFLQVSDVLRCKREYLKIKINAENC 133
52 Db 63 I-CADYDILVILFEFMYTGNVILTKNDNIELVIOQCDVLCIDSLIKICEBYIGIIDEIETNC 121
53 QY 134 VRLISFADLFSCSEELKOSAK-----RMV-----HKFTA----- 162
54 Db 122 IHLNFSFDYTNLQGLRSMKSYRLPKIINNKLTVBELDDMLIIKILKIYIACEYIVYKKI 181
55 QY 163 ----VHQDAFMQSLHLLIDILSSDNL---NVEKEETVEAAMLMLEYNTSSSQYLS 214
56 Db 182 ILNVIIVHKDERIILYTKKLIMKHIIINDQHYTSLSDIELVNNIRERIDYDKHREHDVDDISHPFI 241
57 QY 215 SYLSQ--IRIDALSEVQ-----RAMFQGLRPNDKSVVVOGLYKSNPKPFK 258
58 Db 242 MVGGKIKFNTFAFPLSNKKHIIIDRYDMEGCKTHFSVVYLNLSILYIIIG-----K 292
59 QY 259 PRLQMTKEEMWIFTEASSENPCSLYSVCYSPQAEKVKYLCSPADLHKVGVTVTPND- 317
60 Db 293 KRGFTEKEVLSYNKN-----KLM---CYEP-----LNFPRYDTSVCVSNGM 332
61 QY 318 IYIAGGVPLKNTKTNKTSKLOTARFYVNCFYDAQONTWPRKTPMLFVRIPKSLVC 377

```

Db 333 IYISIGK-----DTNGYMTNIVE-----FM-KPEWKSWMYDGOHLCYPRCWSLVD 376
Qy 378 CEEYIYAIGG--DSVGEINRRTVERD-TEK---DEWTVNSPLPCAMQMSAAVVVHDCI 431
Db 377 YNNEVYTIIGLKTSTIDEFNIEMIIVSDDAVEKLTDSHMKLKQPLAKSGISSIYVNDFI 436
Qy 432 Y-----VMTLNLMYCFPPRSDSVWEMAMRQTSRSPASAAAFGDKIYIGLHI 479
Db 437 YCIGRIDTPHISIEHTNDVYIYSSRDDCKMKYSTNNVARSFCLSCVFNNELIYIGYN- 495
Qy 480 ATNSGIRLPSGTVDSVTVETIDVKNEMKMAANIP 516
Db 496 -TNS-----VEKYNKLTWTKWRLNDIP 516

RESULT 12
US-09-299-268-29
Sequence 29, Application US/09299268
Patent No. 6217882
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Gibbs, E. P. J.
TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSER: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299,268
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/901,127
FILING DATE:
APPLICATION NUMBER: 07/908,241
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 1-JUL-1992
FILING DATE:
APPLICATION NUMBER: US 07/908,630
FILING DATE: 29-JUN-1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/342,212
FILING DATE: 21-APR-1992
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF35.1.FWCCI
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 29:
SEQUENCE CHARACTERISTICS:
LENGTH: 530 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-299-268-29

Query Match 9.4%; Score 309.5; DB 4; Length 530;
Best Local Similarity 21.7%; Pred. No. 3,5e-23;

Matches 125; Conservative 98; Mismatches 215; Indels 139; Gaps 22;
Qy 16 SLLEQLKFEYEQQLFTD--IYLVIGTEPFGKMLATCQSSFRAMFSGSSEKQTVNH 73
Db 3 SLLERLHDFEFGKIGICDIDIKYIENNKTI SARHLTLMYSKTFYNI FNSDFIDKNNDELY 62
Qy 74 LNNVDAATLQIITTAAYGNLAMDSTVEOLYETACFLQVEDVYLQRCREYLKIKNAENC 133
Db 63 I-CADYDILYIILBEMTYGNIVLTQNDIELVQVODYLCIDBLINICEBYIGIIDETNC 121
Qy 134 VLLSPADLFSCBELKQSAK-----RMV-----EHKFTD----- 162
Db 122 IHLNFSPTDYNLQRLREMSKWYLPKIINNKLVELIDIDMILLIKEIKYIACEYIVKKI 181
Qy 163 ----VYHODAFMQLSHDLLIDLSQDNL-----NVEKEETVREAAMLMEYNTESRSQYLS 214
Db 182 IINWIVHDERIITYKKLKHINDODHTLSQDIELYNNIRIYDNKEHDVDISHNFI 241
Qy 215 SVLSQ--IRIDALSEVTO-----RAWFQGLPPNDKSVVVOGLYKSNPKPEK 258
Db 242 MWGKKKIFNITAFNPLSNKKHIIIDRYDMFGCKTHFSVYVLSIYIIG-----K 292
Qy 259 PRLGNTKERMMIFIASSENPCSLYSVCYSPQAEKRYKLCSPPADLHKVGTVPDND- 317
Db 293 KRGYFTKEVLSYNIN-----KLV-----CYEPE-----UNYFRYDTSVCVSNQM 332
Qy 318 IYIAGGOVPLKNTKKNHSTKSLQTAFTVNCFYVPAQOQNTWPKTMLPVRIKPSLYC 377
Db 333 IYISIGK-----DTNGYMTNIVE-----FM-KPEWKSWMYDGOHLCYPRCWSLVD 376
Qy 378 CEEYIYAIGG--DSVGEINRRTVERD-TEK---DEWTVNSPLPCAMQMSAAVVVHDCI 431
Db 377 YNNEVYTIIGLKTSTIDEFNIEMIIVSDDAVEKLTDSHMKLKQPLIASGSIYVNDFI 436
Qy 432 Y-----VMTLNLMYCFPPRSDSVWEMAMRQTSRSPASAAAFGDKIYIGLHI 479
Db 437 YCIGRIDTPHISIEHTNDVYIYSSRDDCKMKYSTNNVARSFCLSCVFNNELIYIGYN- 495
Qy 480 ATNSGIRLPSGTVDSVTVETIDVKNEMKMAANIP 516
Db 496 -TNS-----VEKYNKLTWTKWRLNDIP 516

RESULT 13
US-08-307-499-20
Sequence 20, Application US/08307499
Patent No. 5651972
GENERAL INFORMATION:
APPLICANT: Moyer, Richard W.
APPLICANT: Vi uela, Eladio
TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSER: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307,499
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/908,241
FILING DATE: 1-JUL-1992

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/908,630
FILING DATE: 29-JUN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/342,212
FILING DATE: 21-APR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-307-499-20

Query Match 7.1%; Score 234.5; DB 1; Length 500;
Best Local Similarity 20.4%; Pred. No. 2e-15;
Matches 120; Conservative 92; Mismatches 222; Indels 155; Gaps 25;

1 MSTODROINTEYAVSL-LEQLKLFYEQOLFDTIVLVEGTEPCHQWVLTSSYFRAM 59
1 MSKQ-ETIDYVNIETIRNAVNLRSDDEI--VFIMVGVVKKVKKELVSVSNFYKLI 56
60 FMSGSESQKQTHVLRNVDATLQIITTYATGNLAMDSTYEQLYETACFLQVEDVLR 119
57 TNQGSNEITVSPQESFLD-----IKYIEGTIVTIDLDVNEINISISCSKAIDFLKNS 110
120 CREYLIKKNAENCRVLLSFADLFCSEELKQSAKRWVHKFTAVYHQDAFMOLS-HDLLI 178
111 CIDPMSKAITDSTCVKIKYIKGFSGNGCFAYVNDIAVIRKFTKI-ETDILSLSLFDLRI 169
179 DILSSDNINVEKETVRBAAMLMEYNTESRQYLSVLSQIRIDALS-----EYTORAM 233
170 -ILKSGELDVSEDDVLLFTIKMSRHKKSNNRKSFTLVTVLRVNYLSIGYKLTW--W 226
234 FQGLPNDKSVVVOGLYKSNP-----KFFKPRLCM-----TKEMMIFI 272
227 LARFGKNNN---VELNENELPRISYQHRFTNRKRYMTVPESPSINMLGNVSVNLSIIN 283
273 E-ASSENP-CS-LYSSVCY-----SPQAEKYKLCSPPADLH--KVGIT 310
284 SIAENHNPCGSLVAMDILYLIGINKSLDPVSDITSVDTRSFIEHTPPL-LHPRKCPG 342
311 VTPDNDIYIAGS---QVPLKNTKTHSKTSKQTAFTRYNCCYVWDAQNTMFPRTPL 367
343 VAIETKRIYVGGIGYDGPLKTVES-----WSPGSDQ--RREVPFL 382
368 FVRIKESLVCEBEGYIAYIGDSVGLNRRTERVYDTEKLEMTVMVPLPCAMQMSAAVVV 427
383 QRFNFCIIGTNDILVVGISF---DCKTIEIYEEYRWISIGNAMYSHGCGIAIH 438
428 HDCIYVMTLNLAKCYPRSDSVWEMARQTSRSPASAAAFGDKIFPIYIGLHATNSGIRL 487
439 HGIIYV-----IGGLSFIDNIHV-- 456
488 PGGTVDSGVTVIYDVKNEKMGAMNIPAKRYSDCVRAVVISNLSG 536
457 -----FTWVEKTNPHSKMTVEKSLPEPRF-----NSSLCI 487

RESULT 14
US-09-299-268-20
Sequence 20, Application US/09299268
Patent No. 6217882
GENERAL INFORMATION:

APPLICANT: Moyer, Richard W.
APPLICANT: Vi uela, Eladio
APPLICANT: Gibbs, E.P.J.
TITLE OR INVENTION: Use of Recombinant Swine Poxvirus as a
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: David R. Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299,268
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/901,127
FILING DATE:
APPLICATION NUMBER: US 07/908,241
FILING DATE: 1-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/908,630
FILING DATE: 29-JUN-1992
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/342,212
FILING DATE: 21-APR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-299-268-20

Query Match 7.1%; Score 234.5; DB 4; Length 500;
Best Local Similarity 20.4%; Pred. No. 2e-15;
Matches 120; Conservative 92; Mismatches 222; Indels 155; Gaps 25;

1 MSTODROINTEYAVSL-LEQLKLFYEQOLFDTIVLVEGTEPCHQWVLTSSYFRAM 59
1 MSKQ-ETIDYVNIETIRNAVNLRSDDEI--VFIMVGVVKKVKKELVSVSNFYKLI 56
60 FMSGSESQKQTHVLRNVDATLQIITTYATGNLAMDSTYEQLYETACFLQVEDVLR 119
57 TNQGSNEITVSPQESFLD-----IKYIEGTIVTIDLDVNEINISISCSKAIDFLKNS 110
120 CREYLIKKNAENCRVLLSFADLFCSEELKQSAKRWVHKFTAVYHQDAFMOLS-HDLLI 178
111 CIDPMSKAITDSTCVKIKYIKGFSGNGCFAYVNDIAVIRKFTKI-ETDILSLSLFDLRI 169
179 DILSSDNINVEKETVRBAAMLMEYNTESRQYLSVLSQIRIDALS-----EYTORAM 233
170 -ILKSGELDVSEDDVLLFTIKMSRHKKSNNRKSFTLVTVLRVNYLSIGYKLTW--W 226
234 FQGLPNDKSVVVOGLYKSNP-----KFFKPRLCM-----TKEMMIFI 272
227 LARFGKNNN---VELNENELPRISYQHRFTNRKRYMTVPESPSINMLGNVSVNLSIIN 283

QY 273 E-ASSEN-PCSS-SSVVCY-----SPOAEKYYKLCSPADLH--KVG 310
 Db 284 SIAMNHPYCGSVLMDLILGINKSLDPVDISVDRSPRLTLPPL-LHPRKCPG 342
 QY 311 VTPDNDIYAGC---QVPLKNTKTNHSTKSLQTAFTYTCYVWDQAQNTFPKPTML 367
 Db 343 VAFKRIYVVGIGYDGPLKTVES-----WSPGEQO-WREVPPL 382
 QY 368 FVRIKPSLVCCEGIYVAGDSVSGELNRRITVRDTEKDEMTVSPPLPCAMQMSAAVY 427
 Db 383 QPFPNFIITDNDLIVVGISE---DDKTIETIYEENTWISIGNMANYSHFGCIAH 438
 QY 428 HDCIYVWTLNLMYCYPFRSDSWVEMAMRQTSRSPASAAFGDKIFYIGLHATNSGIRL 487
 Db 439 HGIIYM-----IGLSFIDNIHV-- 456
 QY 488 PSCTVDCSSVTVEIYDVNKNEMKMANIPAKRYSDCPRAVVISNLCV 536
 Db 457 -----FTWVEKYNPHSNKMTVEKSLPFRF-----NSSLCI 487

RESULT 15

US-09-137-223A-2
 / Sequence 2, Application US/09137223A
 / Patent No. 6420525
 / GENERAL INFORMATION:
 / APPLICANT: Yee, David P
 / APPLICANT: Delisher, Theresa A
 / TITLE OF INVENTION: TESTIS-SPECIFIC TRANSCRIPTION FACTOR
 / FILE REFERENCE: ZGCL-1
 / CURRENT APPLICATION NUMBER: US/09/137,223A
 / CURRENT FILING DATE: 1998-08-19
 / PRIOR APPLICATION NUMBER: 06/056,130
 / PRIOR FILING DATE: 1997-08-19
 / NUMBER OF SEQ ID NOS: 16
 / SOFTWARE: FastSeq for Windows Version 3.0
 / SEQ ID NO 2
 / LENGTH: 478
 / TYPE: PRT
 / ORGANISM: homo sapiens
 US-09-137-223A-2

Query Match 5.9%; Score 195.5; DB 4; Length 478;
 Best Local Similarity 21.9%; Pred. No. 2,1e-11;
 Matches 84; Conservative 58; Mismatches 149; Indels 93; Gaps 13;

QY 4 QDERQJNTYAVSLLEQLKLFYEQQLF---TDIVLIVGTEFPCHKVTLATCSSYFRA 58
 Db 41 EQRRLNTPRRKRLKTSKYIY-QTLFLNGENSDIKICALGEWMSLHKIYLCO-SGYFSS 98
 QY 59 MEMSGLSSESQTHVHL-----RNVDAATLOIITVATGNLAMDSTVEQLYETACFLQVE 114
 Db 99 MFGSGWSESMNTIELEIPDONTDEALQVAFGSLYRDVDLIKPSRVVALIAAACLQDD 158
 QY 115 DVLQRCREYLIIKINAENCVRLLSFADLFSCBELKQSAKRMVHKFTAVYHODAFMQLSH 174
 Db 159 GLIQCGETMKETVNVKTVCGYTSAGTYGLDSVKKCKLEWLNLMATHONVELFKELSI 216
 QY 175 DLLIDILSDNLNV-EKETVRBAAMLWL-----BYNTES----- 208
 Db 219 NVMKQLIGSSNLFVMQVEMDIYALKKMPLQLVPSWNGSLKQLLTETDVMFSQKRKPE 278
 QY 209 -----RSQYLSVLSQIRI--DA-----LSEVTQRAWFOGL-P 238
 Db 279 GMAFLETEQCKPVSVFRHLRLQYIISDLASARIIEODAVPSEWLSVYKQGMFAMLR 338
 QY 239 PNDKSYVVOGLYKSMKPFKPRLG--MTXB-----EMMIFIEA 274
 Db 339 EODSEVGPQEIINKELEEGNSMRGCKRLADGEYCMRWGTGFPFDLLVYTRNRIIFKRN 398
 QY 275 SSENPCSLYSSVVCYPOAEKVKYL 298

Db 399 TLNOPS--GSVLOPFRRSIAFRL 420
 Search completed: July 14, 2003, 18:24:28
 Job time : 9.99422 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:16 / Search time 21.5998 Seconds
(without alignments)
3843.329 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3292
Sequence: 1 MSTDQERQINTEYAVSLLEQ.....STDGTEPEFLDGMVALPVV 623

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seque, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3292	100.0	623	22	AAE11886
2	3292	100.0	623	22	AAE11886
3	3204.5	97.3	622	23	AAE60081
4	1675	50.9	335	21	AAE49413
5	578.5	17.6	623	22	AAE42737
6	569.5	17.3	816	22	ABE62185
7	558.5	17.0	609	22	ABG03507
8	555.5	16.9	689	22	AAE92953
9	554	16.8	606	23	ABE63086
10	541.5	16.4	587	22	AAE90399
					AAE38956

11	534	16.2	593	22	AAE78511
12	530	16.1	599	22	AAE79495
13	526.5	16.0	744	22	ABE60307
14	523	15.9	575	23	AAE47944
15	523	15.9	748	22	AAE40225
16	516.5	15.7	589	17	AAE94389
17	512.5	15.6	596	23	ABE60123
18	511.5	15.5	589	17	AAE94386
19	509	15.5	532	21	AAE94936
20	508.5	15.4	584	22	AAE41870
21	505.5	15.4	569	22	ABE11665
22	505.5	15.4	569	22	AAE40204
23	505.5	15.4	569	22	AAE41990
24	505.5	15.4	569	22	AAE94214
25	505.5	15.4	569	23	AAE94302
26	501.5	15.2	521	22	ABE76659
27	501.5	15.2	569	23	AAE79005
28	501	15.2	626	22	AAE97108
29	500	15.2	568	22	AAE94018
30	500	15.2	568	22	AAE94486
31	498	15.1	625	22	ABE59896
32	484.5	14.7	574	22	ABE64408
33	484	14.7	574	22	AAE38711
34	480.5	14.6	575	22	AAE71239
35	478.5	14.5	610	22	AAE40497
36	474	14.4	617	23	ABE97212
37	470	14.3	565	23	ABE81893
38	462	14.0	597	22	AAE93687
39	460.5	14.0	538	22	ABE67078
40	457	13.9	604	22	AAE10610
41	455.5	13.8	584	22	AAE95583
42	455	13.8	538	22	AAE16297
43	453	13.3	506	21	AAE43090
44	435.5	11.3	616	22	AAE95123
45	435.5	11.2	615	22	AAE28187

ALIGNMENTS

RESULT 1	AAE11886	standard; Protein; 623 AA.
ID	AAE11886	
AC	AAE11886;	
DT	18-DEC-2001	(first entry)
DE	Angiogenesis associated kelch-like protein (KLP).	
XX	Angiogenesis associated protein; AAF; cytosolic; cardiac; gene therapy;	
KW	ophthalmological; vulvar; myocardial infarction; macular degeneration;	
KW	diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine;	
KW	rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation;	
KW	cancer; therapeutic; diagnostic; kelch-like protein; KLP.	
XX	Unidentified.	
XX	WO200170808-A2.	
XX	27-SEP-2001.	
XX	22-MAR-2001; 2001WO-US09609.	
XX	22-MAR-2000; 2000US-191134P.	
XX	(CURA-) CURAGEN CORP.	
PA	(GETH) GENENTECH INC.	
XX	Rastelli LK, Gerritsen M;	
PI	WPI, 2001-602775/68.	
DR	N-PsDB; AAD19115.	

XX Novel angiogenesis associated polypeptides and polynucleotides encoding
 PT the polypeptides, useful for modulating angiogenesis and for treating
 PT tumors and cancers -
 XX
 PS Claim 1; Page 9-11, 159pp; English.
 XX
 CC The invention relates to angiogenesis associated proteins (AAP) and their
 CC corresponding cDNA molecules, which are useful for modulating
 CC angiogenesis. AAP proteins and nucleic acids are useful for promoting
 CC wound healing, for example after organ transplantation, and in the
 CC treatment of tumours, myocardial infarction, cancers, diabetic
 CC retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.
 CC AAP proteins and DNA's are useful in potential prophylactic and
 CC therapeutic applications implicated in a variety of disorders including
 CC those related to angiogenesis, and also in diagnostic applications.
 CC AAP cDNA is also useful in gene therapy. The invention also relates to
 CC a method for screening a tissue sample for tumorigenic potential. AAP
 CC proteins are used to screen drugs or compounds that modulate AAP activity
 CC or expression as well as treating disorders characterised by insufficient
 CC or excessive production of AAP or production of AAP forms that have
 CC decreased or aberrant activity compared to the wild type protein, or
 CC modulate biological function that involve AAP. The present sequence is
 CC Kelch-like protein (KLP) which is an angiogenesis associated protein
 CC (AAP) of the invention. KLP is associated with tube formation and
 CC angiogenesis because it is upregulated in the in vitro model of
 CC angiogenesis. Kelch mediates cytoskeletal associations, it is involved in
 CC morphogenetic processes such as tube formation, that depend on
 CC cytoskeletal arrangements and signalling.
 CC
 XX Sequence 623 AA;
 SQ
 Query Match 100.0%; Score 3292; DB 22; Length 623;
 Best Local Similarity 100.0%; Pred. No. 1.4e-297;
 Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSTQDERQINTVEYAVSLLEQLKLFYEQQLFTDIVLIVGTEFPCPKMVLATCSSYFRAMF 60
 DB 1 MSTQDERQINTVEYAVSLLEQLKLFYEQQLFTDIVLIVGTEFPCPKMVLATCSSYFRAMF 60
 QY 61 MSGLSSEKQTHVHLRNVDAAATLQIIITTAATGNLAMDSTVEQLYETACFLQVEDVLQRC 120
 DB 61 MSGLSSEKQTHVHLRNVDAAATLQIIITTAATGNLAMDSTVEQLYETACFLQVEDVLQRC 120
 QY 121 REYLIRKINAENCVRLLSFADLFSCCELKQSAKRWVEHKFTAVYHODA FMQLSHDLLIDI 180
 DB 121 REYLIRKINAENCVRLLSFADLFSCCELKQSAKRWVEHKFTAVYHODA FMQLSHDLLIDI 180
 QY 121 REYLIRKINAENCVRLLSFADLFSCCELKQSAKRWVEHKFTAVYHODA FMQLSHDLLIDI 180
 DB 121 REYLIRKINAENCVRLLSFADLFSCCELKQSAKRWVEHKFTAVYHODA FMQLSHDLLIDI 180
 QY 181 LSSDNLNVEKEFTVREAAMLMLEYNTESRSQYLSVLSQIRIDALSEVTQRAVFOGLPPN 240
 DB 181 LSSDNLNVEKEFTVREAAMLMLEYNTESRSQYLSVLSQIRIDALSEVTQRAVFOGLPPN 240
 QY 241 DKSIVVVOGLYKSNPKPKPRKLGMTKEBMMFIFASSENPSLSVSVCSQAQAKVYKLCG 300
 DB 241 DKSIVVVOGLYKSNPKPKPRKLGMTKEBMMFIFASSENPSLSVSVCSQAQAKVYKLCG 300
 QY 241 DKSIVVVOGLYKSNPKPKPRKLGMTKEBMMFIFASSENPSLSVSVCSQAQAKVYKLCG 300
 DB 241 DKSIVVVOGLYKSNPKPKPRKLGMTKEBMMFIFASSENPSLSVSVCSQAQAKVYKLCG 300
 QY 301 PPADLHVKGTVVPPNDIYIAGGOVPLKTKTNNHSTKSKQTAFFRTVNCYWFDAQONTM 360
 DB 301 PPADLHVKGTVVPPNDIYIAGGOVPLKTKTNNHSTKSKQTAFFRTVNCYWFDAQONTM 360
 QY 301 PPADLHVKGTVVPPNDIYIAGGOVPLKTKTNNHSTKSKQTAFFRTVNCYWFDAQONTM 360
 DB 301 PPADLHVKGTVVPPNDIYIAGGOVPLKTKTNNHSTKSKQTAFFRTVNCYWFDAQONTM 360
 QY 361 FPKTPMLFVRIKPSLVCCBGYIYAIGDSVGGELNRRVRYDTEKDEMTWSPPLPCAMQ 420
 DB 361 FPKTPMLFVRIKPSLVCCBGYIYAIGDSVGGELNRRVRYDTEKDEMTWSPPLPCAMQ 420
 QY 421 WSAVAVVHDCIYMTLNTLMCYFPRSDSWTEAMKQTSNFSASAAAFGDKIFITGGLIHA 480
 DB 421 WSAVAVVHDCIYMTLNTLMCYFPRSDSWTEAMKQTSNFSASAAAFGDKIFITGGLIHA 480
 QY 421 WSAVAVVHDCIYMTLNTLMCYFPRSDSWTEAMKQTSNFSASAAAFGDKIFITGGLIHA 480
 DB 421 WSAVAVVHDCIYMTLNTLMCYFPRSDSWTEAMKQTSNFSASAAAFGDKIFITGGLIHA 480
 QY 481 TNSGIRLPSTGVGSSVTVETIYDVNKNEMMANIPAKRYSDPCVAVYI NSGLCYFMBE 540
 DB 481 TNSGIRLPSTGVGSSVTVETIYDVNKNEMMANIPAKRYSDPCVAVYI NSGLCYFMBE 540
 QY 541 THLNERAKVTVYQYDLELDRWSLRQHSERVLMDLGRDPRCTVKGKLYPSCLEESPMKPT 600

DB 541 THLNERAKVTVYQYDLELDRWSLRQHSERVLMDLGRDPRCTVKGKLYPSCLEESPMKPT 600
 QY 601 YLPSTDGTEEFELDGEWVALPV 623
 DB 601 YLPSTDGTEEFELDGEWVALPV 623
 RESULT 2
 AAB60081
 ID AAB60081 standard; Protein; 623 AA.
 XX
 XX AAB60081;
 AC
 XX
 DT 28-MAR-2001 (first entry)
 XX
 DE Human transport protein TPTP-1.
 XX
 XX Human; transport protein; TPTP; transport disorder; metabolic disorder;
 KW neurological disorder; cardiovascular disorder; reproductive disorder;
 KW immune disorder; cancer.
 XX
 OS Homo sapiens.
 XX
 XX WO200078953-A2.
 PN
 XX
 PD 28-DEC-2000.
 XX
 XX 16-JUN-2000; 2000MO-US16668.
 PF
 XX
 XX 17-JUN-1999; 99US-0139923.
 PR 10-AUG-1999; 99US-0148177.
 PR 18-AUG-1999; 99US-0149357.
 PR 28-OCT-1999; 99US-0162287.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Lal P, Yang H, Hillman JL, Tang YT, Bandman O, Burford N;
 PI Baughn MR, Azimzai Y, Lu DM, Au-Young J, Patterson C;
 DR WPI, 2001-041424/05.
 DR N-PSDB; AAF27701.
 XX
 XX
 PT Isolated polypeptide with a human transport protein sequence is useful
 PT for the diagnosis, prevention and treatment of disorders associated
 PT with the immune, reproductive and cardiovascular systems -
 XX
 XX Claim 2; Page 106-107; 165pp; English.
 PS
 CC The present invention provides the protein and coding sequences for 43
 CC novel human transport proteins (designated TPTPs). These can be used in
 CC the diagnosis and treatment of transport, metabolic, neurological,
 CC reproductive, cardiovascular and immune disorders, and cell proliferative
 CC disorders such as cancer.
 CC
 XX Sequence 623 AA;
 SQ
 Query Match 100.0%; Score 3292; DB 22; Length 623;
 Best Local Similarity 100.0%; Pred. No. 1.4e-297;
 Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSTQDERQINTVEYAVSLLEQLKLFYEQQLFTDIVLIVGTEFPCPKMVLATCSSYFRAMF 60
 DB 1 MSTQDERQINTVEYAVSLLEQLKLFYEQQLFTDIVLIVGTEFPCPKMVLATCSSYFRAMF 60
 QY 61 MSGLSSEKQTHVHLRNVDAAATLQIIITTAATGNLAMDSTVEQLYETACFLQVEDVLQRC 120
 DB 61 MSGLSSEKQTHVHLRNVDAAATLQIIITTAATGNLAMDSTVEQLYETACFLQVEDVLQRC 120
 QY 121 REYLIRKINAENCVRLLSFADLFSCCELKQSAKRWVEHKFTAVYHODA FMQLSHDLLIDI 180
 DB 121 REYLIRKINAENCVRLLSFADLFSCCELKQSAKRWVEHKFTAVYHODA FMQLSHDLLIDI 180
 QY 121 REYLIRKINAENCVRLLSFADLFSCCELKQSAKRWVEHKFTAVYHODA FMQLSHDLLIDI 180
 DB 121 REYLIRKINAENCVRLLSFADLFSCCELKQSAKRWVEHKFTAVYHODA FMQLSHDLLIDI 180

QY 181 LSSDNLNVEKETVEEAAMLMEYNTESRSQYLSVLSQIRIDALSEVTORAMFQGLPNN 240
 DB 181 LSSDNLNVEKETVEEAAMLMEYNTESRSQYLSVLSQIRIDALSEVTORAMFQGLPNN 240
 QY 241 DKSVVVQGLYKSMPEKFKPRLGNTKEEMIFTEASSENPCSLYSVCYSPQAEKYKICS 300
 DB 241 DKSVVVQGLYKSMPEKFKPRLGNTKEEMIFTEASSENPCSLYSVCYSPQAEKYKICS 300
 QY 301 PPADLHKVGTVTTPNDIYIAGGVPLKNTKTNHSTKSLQTFRTVNCFYFMDAQONTM 360
 DB 301 PPADLHKVGTVTTPNDIYIAGGVPLKNTKTNHSTKSLQTFRTVNCFYFMDAQONTM 360
 QY 361 PPKTPEMLFVRIKPSLVCCGGYIYAIAGDSVGGELNRRTYERYDTEKDEWTWVSPLECAMQ 420
 DB 361 PPKTPEMLFVRIKPSLVCCGGYIYAIAGDSVGGELNRRTYERYDTEKDEWTWVSPLECAMQ 420
 QY 421 NSAAVVVHDCIYVMTLNTMYCYFPKSDSWEMAMQTSRSPASAAAFGDKIFYIGGLHIA 480
 DB 421 NSAAVVVHDCIYVMTLNTMYCYFPKSDSWEMAMQTSRSPASAAAFGDKIFYIGGLHIA 480
 QY 481 TNSGIRLPSGTVDSSVTVETIYDVNKNEMKMAANIPAKRISDPCRAVVISNSLCVPMKE 540
 DB 481 TNSGIRLPSGTVDSSVTVETIYDVNKNEMKMAANIPAKRISDPCRAVVISNSLCVPMKE 540
 QY 541 THLNERAKYVTVYQYDLELDRLMSLRQHSERVLMDLGRDFACTVGLYPSCLEESPWKPPPT 600
 DB 541 THLNERAKYVTVYQYDLELDRLMSLRQHSERVLMDLGRDFACTVGLYPSCLEESPWKPPPT 600
 QY 601 YLFSTDTGTEEFELDGEVVALPPV 623
 DB 601 YLFSTDTGTEEFELDGEVVALPPV 623

RESULT 3

AA049413
 ID AA049413 standard; Protein; 622 AA.

AA049413;
 DT 08-OCT-2002 (first entry)

Human mammary gland susceptible gene-encoded protein 68.42.

Human mammary gland susceptible gene-encoded protein 68.42;
 recombinant production; gene therapy; embryonic development disorder;
 tumour; cancer; cytostatic.

Homo sapiens.

CN1340545-A.

20-MAR-2002.

31-AUG-2000; 2000CN-0119828.

31-AUG-2000; 2000CN-0119828.

(BODE-) BODE GENE DEV CO LTD SHANGHAI.

Mao Y, Xie Y;

WPI; 2002-436439/47.

N-PSDB; AB077611.

Polypeptide-human mammary gland susceptible gene coded protein 68.42
 and polynucleotide for coding it -

Claim 1; Page 28-30 (Disclosure); 36pp; Chinese.

The invention relates to human mammary gland susceptible gene-encoded
 protein 68.42 (AA049413) and nucleic acids encoding it (AB077611). The
 protein has a molecular weight of 68.42 kD. The invention also relates to
 a method for the recombinant production of the protein, an antagonist of

CC the protein, and the use of the protein, gene and antagonist in
 CC therapeutic applications. Mammary gland susceptible gene-encoded protein
 CC 68.42 can be used in the treatment of a variety of diseases such as
 CC embryonic development disorders and tumours. The present sequence
 CC represents human mammary gland susceptible gene-encoded protein 68.42.
 XX

Sequence 622 AA:

Query Match 97.3%; Score 3204.5; DB 23; Length 622;
 Best Local Similarity 97.9%; Pred. No. 2e-289;
 Matches 610; Conservative 2; Mismatches 10; Indels 1; Gaps 1;

QY 1 MSTODEROINTEAVASLLEBOLKLFYEQQLFTDYLIVETGEPPCHKVYLATSSYFRAMF 60
 DB 1 MSTODEROINTEAVASLLEBOLKLFYEQQLFTDYLIVETGEPPCHKVYLATSSYFRAMF 60
 QY 61 MSGLESKQTHVHLNNDAAATQIITTAAYGNLAMDSTVEQLYETACFLQVEDVLQRC 120
 DB 61 MSGLESKQTHVHLNNDAAATQIITTAAYGNLAMDSTVEQLYETACFLQVEDVLQRC 120
 QY 121 REYLKKINAENCVRLLSPADLFSCCELKQSAKRVKHTFVAVYQDAFMQLSHLLDI 180
 DB 121 REYLKKINAENCVRLLSPADLFSCCELKQSAKRVKHTFVAVYQDAFMQLSHLLDI 180
 QY 181 LSSDNLNVEKETVEEAAMLMEYNTESRSQYLSVLSQIRIDALSEVTORAMFQGLPNN 240
 DB 181 LSSDNLNVEKETVEEAAMLMEYNTESRSQYLSVLSQIRIDALSEVTORAMFQGLPNN 240
 QY 241 DKSVVVQGLYKSMPEKFKPRLGNTKEEMIFTEASSENPCSLYSVCYSPQAEKYKICS 300
 DB 241 DKSVVVQGLYKSMPEKFKPRLGNTKEEMIFTEASSENPCSLYSVCYSPQAEKYKICS 300
 QY 301 PPADLHKVGTVTTPNDIYIAGGVPLKNTKTNHSTKSLQTFRTVNCFYFMDAQONTM 360
 DB 301 PPADLHKVGTVTTPNDIYIAGGVPLKNTKTNHSTKSLQTFRTVNCFYFMDAQONTM 360
 QY 361 PPKTPEMLFVRIKPSLVCCGGYIYAIAGDSVGGELNRRTYERYDTEKDEWTWVSPLECAMQ 420
 DB 361 PPKTPEMLFVRIKPSLVCCGGYIYAIAGDSVGGELNRRTYERYDTEKDEWTWVSPLECAMQ 420
 QY 421 NSAAVVVHDCIYVMTLNTMYCYFPKSDSWEMAMQTSRSPASAAAFGDKIFYIGGLHIA 480
 DB 421 NSAAVVVHDCIYVMTLNTMYCYFPKSDSWEMAMQTSRSPASAAAFGDKIFYIGGLHIA 480
 QY 481 TNSGIRLPSGTVDSSVTVETIYDVNKNEMKMAANIPAKRISDPCRAVVISNSLCVPMKE 540
 DB 481 TNSGIRLPSGTVDSSVTVETIYDVNKNEMKMAANIPAKRISDPCRAVVISNSLCVPMKE 540
 QY 541 THLNERAKYVTVYQYDLELDRLMSLRQHSERVLMDLGRDFACTVGLYPSCLEESPWKPPPT 600
 DB 541 THLNERAKYVTVYQYDLELDRLMSLRQHSERVLMDLGRDFACTVGLYPSCLEESPWKPPPT 600
 QY 601 YLFSTDTGTEEFELDGEVVALPPV 623
 DB 601 YLFSTDTGTEEFELDGEVVALPPV 623

RESULT 4

AA042737
 ID AA042737 standard; Protein; 335 AA.

AA042737;
 DT 08-FEB-2001 (first entry)

Human ORFX ORF2501 polypeptide sequence SEQ ID NO:5002.

Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnerability; antiproliferative; antiparkinsonian; neuroprotective;
 KW anticonvulsant; osteopathic; antidiabetic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasodilator; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; anti-inflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antithyroid;

KM anti-naemic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage disease; anti-inflammatory disease; coagulation;
 KM thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 PN MO20005473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinketsu RA, Leach M;
 XX
 DR WPI: 2000-602362/57.
 DR N-PSDB; AAC76946.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancer, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 4178-4179; 5507pp; English.
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antarthritic; immunosuppressant;
 CC immunostimulant; cardiac; thrombolytic; coagulant; vasoactive;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC anti-inflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypertoid; anti-naemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancer,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, anti-inflammatory disease, to enhance
 CC coagulation, to inhibit thrombosis, and as a contraceptive.
 XX
 XX Sequence 335 AA:
 SQ
 Query Match 50.9%; Score 1675; DB 21; Length 335;
 Best Local Similarity 98.8%; Pred. No. 2.8e-147;
 Matches 337; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MSTDQROINTFAVSLLEQLKFFEQQLFTDIVLVEGTFPCCHQVATCSYFRANF 60
 DB 1 MSTDQROINTFAVSLLEQLKFFEQQLFTDIVLVEGTFPCCHQVATCSYFRANF 60
 QY 61 MSGLSKQTHVLRNVDAATLQIITVAYTGNLANNDSTVEQVETACGLQVEDVLCRC 120
 DB 61 MSGLSKQTHVLRNVDAATLQIITVAYTGNLANNDSTVEQVETACGLQVEDVLCRC 120
 QY 121 REYLKIKINAENCVRLLSPADLFSCEBLKQSAKRWVHKFTAVYHODAFMQLSHDLLIDI 180
 DB 121 REYLKIKINAENCVRLLSPADLFSCEBLKQSAKRWVHKFTAVYHODAFMQLSHDLLIDI 180

DB 121 REYLKIKINAENCVRLLSPADLFSCEBLKQSAKRWVHKFTAVYHODAFMQLSHDLLIDI 180
 QY 181 LSSDNLNVEKERTVEAAMLMEYNTESRSOVLSSQIRIDALSEVTORAMFQGLPN 240
 DB 181 LSSDNLNVEKERTVEAAMLMEYNTESRSOVLSSQIRIDALSEVTORAMFQGLPN 240
 QY 241 DKSVVVOGLYKSMFPFKRLMTKEBMMIFIEASSENPCSLYSSVCYSPQAEKYKLC 300
 DB 241 DKSVVVOGLYKSMFPFKRLMTKEBMMIFIEASSENPCSLYSSVCYSPQAEKYKLC 300
 QY 301 PPADLHKGTGTVTPNDIYIAGGVPLKRTX 331
 DB 301 PPADLHKGTGTVTPNDIYIAGGVPLKRTX 331
 RESULT 5
 ID ABB62185 standard; Protein; 623 AA.
 XX
 AC ABB62185;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster polypeptide SEQ ID NO 13347.
 XX
 KM Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PMD, Myers EW;
 XX
 DR WPI: 2001-656860/75.
 DR N-PSDB; ABL06288.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Disclosure; SEQ ID NO 13347; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABBS7737-ABR72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 623 AA:
 SQ
 Query Match 17.6%; Score 578.5; DB 22; Length 623;
 Best Local Similarity 27.4%; Pred. No. 1.3e-44;
 Matches 155; Conservative 103; Mismatches 215; Indels 93; Gaps 15;
 QY 18 LEQLKFFEQQLFTDIVLVEGTFPCCHQVATCSYFRANFMSGLSKQTHVLRNV 77
 DB 59 LTELMLTRHRLCDVLNVGGRKIFAHRVILSACSSYFCAMFTGLBESRQTEVTIRDI 118

XX	AA92953	
XX	ID	AA92953 standard; Protein: 609 AA.
XX	AC	AA92953;
XX	AA92953;	
XX	DT	26-JUN-2001 (first entry)
XX	DE	Human protein sequence SEQ ID NO:11635.
XX	Human	protein sequence SEQ ID NO:11635.
XX	Human,	primer: detection; diagnosis; antisense therapy; gene therapy.
XX	OS	Homo sapiens.
XX	PN	EPI074617-A2.
XX	PD	07-FEB-2001.
XX	PF	28-JUL-2000; 2000EP-0116126.
XX	PR	29-JUL-1999; 99JP-0248036.
XX	PR	27-AUG-1999; 99JP-0300253.
XX	PR	11-JAN-2000; 2000JP-0118776.
XX	PR	02-MAY-2000; 2000JP-0183767.
XX	PR	09-JUN-2000; 2000JP-0241899.
XX	PA	(HELI-) HELIX RES INST.
XX	XX	
XX	P1	Ota T, Isogai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;
XX	P1	Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	XX	WPI, 2001-318749/34.
XX	PT	Primer sets for synthesizing polynucleotides, particularly the 5602
XX	PT	full-length cDNAs defined in the specification, and for the detection
XX	PT	and/or diagnosis of the abnormality of the proteins encoded by the
XX	PT	full-length cDNAs -
XX	PS	Claim 8; SEQ ID 11635; 2537bp + CD ROM; English.
XX	PS	
XX	XX	The present invention describes primer sets for synthesizing 5602
XX	XX	full-length cDNAs defined in the specification. Where a primer set
XX	XX	comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX	XX	to the complementary strand of a polynucleotide which comprises one of
XX	XX	the 5602 nucleotide sequences defined in the specification, where the
XX	XX	oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX	XX	of an oligonucleotide comprising a sequence complementary to the
XX	XX	complementary strand of a polynucleotide which comprises a 5'-end
XX	XX	sequence and an oligonucleotide comprising a sequence complementary to a
XX	XX	polynucleotide which comprises a 3'-end sequence, where the
XX	XX	oligonucleotide comprises at least 15 nucleotides and the combination of
XX	XX	the 5'-end sequence/3'-end sequence is selected from those defined in
XX	XX	the specification. The primer sets can be used in antisense therapy and
XX	XX	in gene therapy. The primers are useful for synthesizing polynucleotides,
XX	XX	particularly full-length cDNAs. The primers are also useful for the
XX	XX	detection and/or diagnosis of the abnormality of the proteins encoded by
XX	XX	the full-length cDNAs. The primers allow obtaining of the full-length
XX	XX	cDNAs easily without any specialised methods. AAH03166 to AAH1628 and
XX	XX	AAH13533 to AAH18742 represent human cDNA sequences; AA92446 to
XX	XX	AA95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX	XX	represent oligonucleotides, all of which are used in the exemplification
XX	XX	of the present invention.
XX	XX	
XX	Sequence	609 AA;
XX	XX	

	Query Match	Similarity	Score	DB 22	Length	609
Best Local	150	Conservative	110	Mismatches	222	Gaps
Matches	150	Conservative	110	Mismatches	222	Gaps
9	INTERAVALS	LBQKLFYEOLFTDYLVEGEGFCHKAVL	ATC	SSYFPAM	SGSLSEK	68
46	ISDKKPROTEIN	LRKRRKRELCVGLVGA	AKTIAH	RIYLSAC	SYFPAM	TGELABSR
69	OTHVHLNRVADA	TLOIIITVAATGNL	AMDSVFEOL	YETACELQVEDV	YQRCREYL	IKKI

```

Db 106 QTEVAVIRDDERAGAMELLIDFATTSQITVEBGNVQTLPRACLLQJLAIQACCEFLKROL 165
Qy 129 MAENCYRLLSPADLPSCEELKOSAKRWYEHKFTAVYHODAFMQLSHDLLIDILSSDNLV 188
Db 166 DPNCGGIRAFADTHSCRELLRIADKQFHNFQJEWSEEFMLLPANQLIDISSDELAV 225
Qy 189 EKEETREAMLMLENTESRSQYSSLSUQIRIDALS-----E 227
Db 226 RSEQGFANVMAWVKISIOERRPQJPOVLQHRLELSPKLVGTGSDPLISDSECRD 285
Qy 228 VTQRAVFOGLPRNDGSVVVQGLYKSMPEFKPRJLGMTKEEMPIEASSENPCS---LYS 284
Db 286 LVDEAKNYVLLPQBER-LMQG-----PR-TPRKIRGSEVL---SAGWGSCGDAISN 334
Qy 285 SVCYSQAEKVYKLGSPPADLHKVGTVTVPNDIYIAG-----QVPLKTKTNH-- 334
Db 335 VERYPDQJENBEMVMVMSMRKRCGVSVYL-DLLLYAVGHDGSSYLSVRYDPKTKQMS 393
Qy 335 ---SKTSKQTA-----FRYNCFWPFPAQONTMFPKTPMLPVYIKP 373
Db 394 SDMAPISTCRITSQCAVALGGLYAVGGDGVSCNLIVERYDPKKNKTRVASHSTRRLGV 453
Qy 374 SLVCCGXYIYAIGSDSVGGELNRRFVERDYDEKDMWTSPLPCAMOMSAVWHDCIYV 433
Db 454 AVAVLGGFLYAVGSDGTSPLN--IVERYYNQERMHMTIAMGRRRHGLGAYQDMIVA 511
Qy 434 M-----TLNLMYCYFPRSDSVWEMAMKQTSRSPASAAAFGDKIFITGGHIAHNSGIR 466
Db 512 VGGRDDTTEISSLAERNPRTNQMSFVAVMTSRSGVLAVVAGOLMAVGGF----- 562
Qy 487 LPBGTVDGSSV--TVEIYVNVKNKEMAMNIPAKR 519
Db 563 -----DGTYYLKTIEVEPDNATKRLGGMATYKR 591

```

RESULT 8
ABBB63086
ID ABBB63086 standard; Protein; 689 AA.
XX
AC ABBB63086;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 16050.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
pharmaceutical.
XX
OS Drosophila melanogaster.
XX
MO200171042-A2.
XX
PN
XX
PD 27-SEP-2001.
XX
PE 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656660/75.
DR N-PSDB; ABL07189.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 16050; 21pp + Sequence Listing; English.
XX

CC The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from *Drosophila*. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 689 AA;

Query Match 16.9%; Score 555.5; DB 22; Length 689;

Best Local Similarity 27.1%; Pred. No. 2,1e-42;

Matches 168; Conservative 104; Mismatches 246; Indels 103; Gaps 22;

QY 10 NTEVAVSLIEQLKLFYEQQLFTDIVLIVEGTEPCHRMVLAATCSYFRAMFSGLSKQ 69
 DB 136 NEQHTARSPDAMNEMKQKQLCVILVADVEIHAHMYLASCSPYFAMFTS-FRESQ 194
 QY 70 THVHLRNVDATLQIIITAYTGNLAMDSTVEQLYETAKFLOVEDVLRCHREYLKIKIN 129
 DB 195 ARITLQSVARALELIDVYATATVENVNDNVQLLTAAVLQTLTVRDACDFLOTOLD 254
 QY 130 AENCVALLSFADLFSCCEELKQSAKRMVEHFTAVYHQAAMQSLHDLIDLSQNLAYE 189
 DB 255 AENCGLIRFADIHACVELLNYAETIEQHFNVEIOFDEFLNLSHQVSLIGNDLSIP 314
 QY 190 KEETVREAAMLLEVNTESRSQYLSVLSQIRIDALSE--VTQRAMFQGLPNDKSVVQ 247
 DB 315 NEEVRYECVIAVLRVDPMEQFTSLMEHVRLLPFLSKETIQRV-----DKETLLE 366
 QY 248 G-----LYSMKPFK-----PR--LGMKEMMFIEMSSNPGSLYSVCY 288
 DB 367 GNIVCKNLIIEALTYHLPLETKSARTVPRKPKVMK---ILLVIGQAPKAIRSEVNY 422
 QY 289 SPQAEKYVYLCSPPADLHKVGVTVTPDNDIYIAGGVPLNKTNTNHSKTSKLTQAPRTYN 348
 DB 423 DIREKMYQAAEMPRNCRSGSLVGDK-VYAVG-----FNCS-----LKRITVD 467
 QY 349 CFYWFDAQNTWFPKTPMLFVRIKPSLVCCGYIYVAGISVSGELNRRTERVETKDE 408
 DB 468 V---VDPATDQWANCNMEARRSTLGVAVLNGCIYAVG--FDGTTGLSAAEWDPKTDI 522
 QY 409 WTMVSPLPQAMQWAAVYVHDCIYV-----TLNLMKYCFPSPDSWEMAMQTR 459
 DB 523 WRFIASMSTRSSVGVVHGLLYAVGVDGFTROCLSSVERNPDTDTMVVVAEMSSRR 582
 QY 460 SPASAAAFGDKIFYIGLHATNSGIRLPSTGTVDGSSV--TVEIYVNNKEMMANIPA 517
 DB 583 SGAGVGVNNTIYAVGG-H-----DGMVRRSEAYICETNSMSVADMSY 627
 QY 518 KYSDEPCVAVVISNS--LCVFMRETHLNERAKYVTYQVLELDMSLRQHSERVLMDL 575
 DB 628 CRRN-----AGVVAHGDGLYVVGDDGTGNSLASEVYCPD--SDSW-----RLPAL 672
 QY 576 GRDFCTVGLTPS-CLESP 595
 DB 673 -----WTIGRSYAVGCMIDKP 688

RESULT 9

ABB90399

ID ABB90399 standard; Proteoin; 606 AA.

XX ABB90399;

XX 24-MAY-2002 (first entry)

DE Human polypeptide SEQ ID NO 2775.

KW Cytostatic; immunosuppressive; nocitropic; neuroprotective; antiviral; antiallergic; hepatotropic; antidiabetic; antiinflammatory; antitumor; antileukemic; anticonvulsant; antibacterial; antifungal; antiparasitic; KX cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder; KW neurological disease; infection; human; secreted protein.

OS Homo sapiens.

PN WO200190304-A2.

XX 29-NOV-2001.

PF 18-MAY-2001; 2001WO-US16450.

PR 19-MAY-2000; 2000US-205515P.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Birse CE, Rosen CA;

DR WPI; 2002-122018/16.

XX N-PSDB; ABL90808.

PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and prevention of neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal and proliferative disorders -

PS Claim 11; SEQ ID NO 2775; 2081bp + Sequence Listing; English.

CC The invention relates to novel genes (ABL89449-ABL90853) and proteins (ABB89040-ABB90444) useful for preventing, treating or ameliorating medical conditions e.g. by protein or gene therapy. The genes are isolated from a range of human tissues disclosed in the specification. CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and ovarian cancer and other cancers of the adrenal gland, bone, bone marrow, breast, gastrointestinal tract, liver, lung, or urogenital; CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c) cardiovascular disorders such as myocardial ischaemia; CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f) infectious diseases such as viral, bacterial, fungal CC and parasitic infections.

CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 606 AA;

Query Match 16.8%; Score 554; DB 23; Length 606;

Best Local Similarity 26.6%; Pred. No. 2.4e-42;

Matches 155; Conservative 93; Mismatches 240; Indels 94; Gaps 13;

QY 6 BROINTE---YAVSLIEQ-LKLFYEQQLFTDIVLIVEGTEPCHRMVLAATCSYFRAMF 61
 DB 4 QRELAELRLVSTLQDGLDLDLDBKFTICTLAKAGSKSPCHNLIISACSPYRBYTL 63
 QY 62 SGLSEKQTHVLRNVDAATLQIIITAYTGNLAMDSTVEQLYETAKFLOVEDVLRCH 121
 DB 64 SVIDAKKEVVDVNDPAILDLILIKYLSASIDLDNDGNVDIPLASRFPQIPSYFVGV 123
 QY 122 EYLKIKIAENGVRLLSFADLFSCCEELKQSAKRMVEHFTAVYHQAAMQSLHDLIDL 181
 DB 124 STLQRLAPGNCALIRGLLDCPRLAISAEFVSDRFVQCKEDFPQLSPQELISVI 183
 QY 182 SSDNLNVEKETVREAAMLLEVNTESRSQYLSVLSQIRIDALSE----- 227
 DB 184 SNDSLNVEKEAVFPAVWKVTRTDKENVKSLSVFDICIRPLMTEKKPKHVEKODI 243
 QY 228 -----VTQRAMFQGLPNDKSVVVGGLYK-----SNPKPFK--PRIGMTKE 266

Db 244 SNEPDLQKIKYLVKDAFAKGLPEBPKNAKTGAGEVNGDVEDLLPGYLVNDIPRHGNEVK 303
Qy 267 EMMIFIBASSENPCSLYSVCSYSPQAEKVVYKLCSPADLHKVGTVPDNDIYAGQVP 326
Db 304 DLILLVNDT-----AAVAYDTENECLTALAEIPRNHSGIYVQOQIYVGGLY- 354
Qy 327 LKQTKTNHSTSKLQTAFTRTVNCYWFDAQONTWFPKTPMLFVRIKPSLVCCRGY----- 361
Db 355 -----VDEENKQDPLQ-----SYFFQDLSIASEWVGPLP-----PSARCLFGLGEVDD 398
Qy 382 -IYAIGSDSVGELNRTVRYDEKDEMTVSPKPCAMQMSAAVVVHDDIYV----- 434
Db 399 KIYVAVACKDQTEASLDSVLCYDPAVAAKMEVKKLPIKXVGHNVISHKMTYCLGKTTD 458
Qy 435 --TLNLMYCYFPSPDSWVENAMRQTSRSPASAAFGDKIFYIGLHATNSGIRLPSGTV 492
Db 459 KKCTNRVFIINPKKGDMDKDLAPMKIPRSMFGVAVHKKIYIAGV-----TE 505
Qy 493 DGSSTVEIYDVNKNEMKMANIPAKRYSDPCYRAVVINSNL 534
Db 506 DGLSASVEAFDLITTNKMDVMTFEPQERS--ISLVSLAGSL 544

RESULT 10

AA038956 standard; Protein; 587 AA.

AA038956;

22-OCT-2001 (first entry)

Human polypeptide SEQ ID NO 2101.

Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager Syndrome; Chemotactic; leukemic; thrombolytic; drug screening; arthritis; inflammation;

Homo sapiens.

MO20015312-A1.

26-JUL-2001.

26-DEC-2000; 2000MO-US34263.

21-JAN-2000; 2000US-0488725.

25-APR-2000; 2000US-0552317.

09-JUL-2000; 2000US-0598042.

19-JUL-2000; 2000US-0620312.

03-AUG-2000; 2000US-0653450.

14-SEP-2000; 2000US-0662181.

19-OCT-2000; 2000US-0693036.

29-NOV-2000; 2000US-0727344.

(HYSE-) HYSEQ INC.

WPI; 2001-442253/47.

N-PSDB; AAI58112.

Novel nucleic acids and polypeptides, useful for treating disorders

such as central nervous system injuries -

The invention relates to human nucleic acids (AA157798-AA161369) and the encoded polypeptides (AA038642-AA042213) with noctropic,

CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localized neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilization of the activities such as: Immune system suppression,
CC actin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemia and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.

Sequence 587 AA;

Query Match 16.4%; Score 541.5; DB 22; Length 587;

Best Local Similarity 25.9%; Pred. No. 3.3e-41;

Matches 156; Conservative 106; Mismatches 216; Indels 125; Gaps 19;

Qy 14 AVSLLEQLKFEQOLFDTIVLVGTEPPCHQWVLAATCSYFRAMFGSLSESKOTVH 73
Db 36 AFKVMNEER--SKQLLQDWTVAEDVEIAHRVLAACSPYFCAMFTGDMSESKAKTE 92
Qy 74 LNNVDAATLQIITTYAATGNLAMDSTVEOLYETACFLOVEDVLCRCREYLKIKINAENC 133
Db 93 IKDVOGQLSKLIDYITTAIEVTEENQVLLPASLSLQMDVRQNCDFLOSLHPTNC 152
Qy 134 VRLISFADLPSCBELKQSAKRVNHFVTAHYQDAFMQSLHLLIDILSDNLNVEKET 193
Db 153 LGIRAFADVHTCTDLQCANAAAEQHFPEVMGSEBFLSLDQVCSLSSDRLTVSSBEK 212
Qy 194 VBEAMLMLEYTESRSQVLSQIRDALSEVTQRAWFGGLPNDSSVVVQGLYKSM 253
Db 213 VBEAVISWVNYKENLEHMAKLEHVRPL-----PRD--YVQ----- 251
Qy 254 PRFPRKRLQMTKEEMVPIEASSENPCSLYSVCSYSPQAEKVVYKLCSPADLH----- 306
Db 252 -----TVEEHALI-----KNNNTCK-----DFLEAMKYHL-----PLDQRLLIKNDP 290
Qy 307 -KVGIVVTPDNDIYIAGQVPLKNTKTNHSTSKLQTAFTRTVNCYWFDAQONTWFPKTP 365
Db 291 TKRTPFVSLPKVMIVVGGQAP-----KIRSVCC--YDFEEDRMDQIAE 332
Qy 366 MLFVRIKPSLVCCGTYAIGSDSVGELNRTVRYDEKDEMTVSPKPCAMQMSAAV 425
Db 333 LPSRRCRAGVFMAGVITAVG--FNGSLRVRTVYDGVKQMTSISMQERRSTLGA 390
Qy 426 VVHDCIYVW-----TLNLMYCYFPSPDSWVENAMRQTSRSPASAAFGDKIFYIG-- 476
Db 391 VANDLLYAVGFDGSLASVEAYGKTNEMFVAPMNTROS VGVVGGGLYAVGVD 450
Qy 477 -----LHIA-----TNSGIRLPSGTV-----DGSV--TVEIY 502
Db 451 GASROCLSTEVEQYNPATNEMIVYADMTSRSGAGVSGQLYATGSHGDLVRSVEVY 510
Qy 503 DVNKNEMKMANIPAKRYSDPCVR--AVVISLQVPMRETHLNRKAVVTVQVYLEDR 560
Db 511 DPGTNTKQVADM-----NMCRRNAGCAVGLLVGGDDGSCNLASVEY--YNPVTDK 563
Qy 561 WSL 563
Db 564 WTL 566

RESULT 11

AA078511 standard; Protein; 593 AA.

AA078511;

06-NOV-2001 (first entry)

XX Human protein SEQ ID NO 1173.
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX Human; cytokine; cell proliferation; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 OS Homo sapiens.
 XX MO200157190-A2.
 XX PD 09-AUG-2001.
 XX PF 05-FEB-2001; 2001WO-US04098.
 XX PR 03-FEB-2000; 2000US-0496914.
 XX PR 27-APR-2000; 2000US-0560875.
 XX PR 20-JUN-2000; 2000US-0598075.
 XX PR 19-JUL-2000; 2000US-0620325.
 XX PR 01-SEP-2000; 2000US-0654936.
 XX PR 15-SEP-2000; 2000US-0663561.
 XX PR 20-OCT-2000; 2000US-0693325.
 XX PR 30-NOV-2000; 2000US-0728422.
 XX (HYSE-) HYSEQ INC.
 XX PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
 PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 XX DR N-PSDB; AAKS1644.
 XX PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX Claim 20; Page 3416-3417; 6221pp; English.
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAK60302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various immunomodulatory activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM60020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX SQ Sequence 593 AA;
 Query Match 16.2%; Score 534; DB 22; Length 593;
 Best Local Similarity 25.3%; Pred. No. 1,7e-40;
 Matches 138; Conservative 100; Mismatches 199; Indels 108; Gaps 14;
 QY 5 DERQINTE-----YAVSLLEQLKLFYEQQLFTDVLIVEGTEFPCHKVLAAT 51
 DB 20 DSKDNTEHGCVTVVPMHMKKAFKVMNELR--SQNLICDVTIYVEMDEISHRVLLAA 76
 QY 52 GSSTPRAMMSGLSEKQTHVHLKNDATLTQIIITVYATGNLANNDSTVEQLYETACPL 111
 DB 77 CSPTPRAMPTGMSSESRARVRKEVDGWTLRMLDYVTAIOVTEENVQVLLPAAAGLL 136
 QY 112 QVEDVLPQREYLIKIKINACVRLSPADLPSCEELTKQSAKRMVHKFTAYVHODAFMQ 171
 DB 137 QODVYKTCERFLESQHLHVNCLGIRAFADMACTDLMLKANTYAEQHFADVVLSEEFIN 196
 QY 172 LSHDLLIDILSSDNLNVEKETVREAAAMLMEYTESRQQLSVLSQIRIDALSB--VT 229

DB 197 LGIEQVCSLISDSKLTLSSEKVFPAVIAVWHDVQEFMARIMHVRLLPLPREYLV 256
 QY 230 ORAMWQGLPND---KSVVQGL--YKSMKPFKRLMTGKEMMIPIBASNNPCLSVSS 285
 DB 257 QVEEBEALVKKSSACKNLYLEMKYHLIP-----TEQKILM----- 292
 QY 286 VCYSPOAEKVYLLCSPPADLHKGVTVPDNDIYAGQVPLKNTKTNSTKSLQTAFR 345
 DB 293 -----KSVTRLRLTPMNLPLKLMVV-----GGOAP-----KAIR 321
 QY 346 TVNCFYWFDAQONTWPKTPMLFVRIKPSLVCCBGYIYVAGDSVGBLNRRTRYDTE 405
 DB 322 SVEC---YDFKEQRHQAELPSRRCRAGMYMAGLVFAVG3--FNJSLRVTVDSYDPV 376
 QY 406 KDEMTMVSPLPCAMQMSAAVVHDCIYV-----TLNLMCYFPRSDSWEMARQTS 458
 DB 377 KDQMTSVAMNRDRRSTLGAALVINGLLVAVGGFDGSGTSSVEAYVIKSNEMFVAPMNR 436
 QY 459 RSFASAAAFGDKIFYIGLHATNSGIRLPSTVDGSS-----VVEIYDVNKNEMKMAN 514
 DB 437 RSSVGVGWGLLVAVG3-----YDGA8ROYLSTVECYNATTNENTYIAE 481
 QY 515 IPAKR 519
 DB 482 MSTRR 486
 RESULT 12
 AAM79495
 ID AAM79495 standard; Protein; 599 AA.
 XX AAM79495;
 AC AAM79495;
 XX DT 06-NOV-2001 (first entry)
 XX DE Human protein SEQ ID NO 3141.
 XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorder; arthritis; inflammation.
 OS Homo sapiens.
 XX MO200157190-A2.
 XX PD 09-AUG-2001;
 XX PF 05-FEB-2001; 2001WO-US04098.
 XX PR 03-FEB-2000; 2000US-0496914.
 XX PR 27-APR-2000; 2000US-0560875.
 XX PR 20-JUN-2000; 2000US-0598075.
 XX PR 19-JUL-2000; 2000US-0620325.
 XX PR 01-SEP-2000; 2000US-0654936.
 XX PR 15-SEP-2000; 2000US-0663561.
 XX PR 20-OCT-2000; 2000US-0693325.
 XX PR 30-NOV-2000; 2000US-0728422.
 XX (HYSE-) HYSEQ INC.
 XX PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
 PI Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
 XX WPI; 2001-476283/51.
 XX DR N-PSDB; AAKS2628.
 XX PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 XX Claim 20; Page 264; 6221pp; English.

XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78323-AAM60302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activity/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM60020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 XX
 XX Sequence 599 AA;
 Query Match 16.1%; Score 530; DB 22; Length 599;
 Best Local Similarity 25.1%; Pred. No. 4e-40;
 Matches 137; Conservative 100; Mismatches 200; Indels 108; Gaps 14;
 QY 5 DEQINTP-----YAVSLLEQLKLFYEQQLFTDVLIVGTEPCHKAVLAT 51
 Db 26 DSKDDTEKHCPTVTPMWMKKAFKYNNELR---SQNLCDVTIVADEMISAHRYVLA 82
 QY CSSEYFRAMFMSGLSESQKOTVHNLRVNDAATLQIIITAYTGNLAMDSTVEQLYETACFL 111
 Db 83 CSSEYFRAMFMSGLSESBAKVRKKEVDGMLRLIDVYTAELQVTEENVQLLPPAGLI 142
 QY 112 QVEDVLORCREYLIIKKINAEVCVRLLSFADLFSCBELKQSAKRMVHKFTAVYHODAFMO 171
 Db 143 QLDGVKTCCEFLSEQLHPNCGIRAFADMACTDLINKANTYAAQHFAVDVLSEEFIN 202
 QY 172 LSHDLIDILISSNINAVEKETREAAMLLENTSRSQYSSVLSQIRIDALSE--VT 229
 Db 203 LGIEQVCSLISSDKLTISSEKVEAVIAVMNNDKVOEFMARLMEHVRLLPLPREYLV 262
 QY 230 QRAMFQGLPND---KSVVVOGL-YKSMPEFKRLGTMTEEMWIFLEASSENPCSLYS 285
 Db 263 QRVEERLVNNSACKNYLLEAKKTHLP-----TEQRILM----- 298
 QY 286 VCYSPOAEKYYKLCSPADLHKVGTVPNDIYIAGGVPLNKTNTNHSKTSKLOTAPR 345
 Db 299 -----KSVRTLRTPMNLKPLMVVY-----GQGNP-----KALR 327
 QY 346 TVNCFWFPDAQONTWPCKTMLFVRIKPSLYCEGCIYIAGDSVGGELNRRPTVERKYDE 405
 Db 328 SAEC---YDFKEGRHQVAFELPSRRCRAGNYLAGLVFAVG--FNGSLRVRITVDSYDPV 382
 QY 406 KDEMTWVSLPLPCAMQMSAAVVVHDCIYVM-----TLNLMCYCFPRSDSVWEMAROTS 458
 Db 383 KQOMTSVANMRDRSRSLGAANVINGLIVAGGFDGSGISSEVAYNINKSNEMFVAAPMNR 442
 QY 459 RSFASAAAFGDKI FYIGGLHIAITNSGIRLPSGTVDSS-----VTVEIYDVNKNEMWMAAN 514
 Db 443 RSSVGVGVGGLIYAVG-----YDGNASROYLSTVCYNAATNTMETWYIAE 487
 QY 515 IPAKR 519
 Db 488 MSTRR 492

RESULT 13

ID ABB60307 standard; Protein; 744 AA.

XX ABB60307;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 7713.

KW Drosophila; developmental biology; cell signalling; insecticide;
 KM pharmaceutical.
 XX Drosophila melanogaster.
 OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX (PEKE) PE CORP NY.
 PA Venter JC, Adams M, Li PMD, Myers EW,
 PI WPI; 2001-656660/75.
 DR N-PSDB; ABL04410.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 XX Disclosure; SEQ ID NO 7713; 21pp + Sequence Listing; English.
 PS
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL01940-ABL6175) and the encoded proteins
 CC (ABBS7737-ABBS72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 744 AA;
 Query Match 16.0%; Score 526.5; DB 22; Length 744;
 Best Local Similarity 28.0%; Pred. No. 1.2e-39;
 Matches 157; Conservative 77; Mismatches 207; Indels 119; Gaps 18;
 QY 11 TEYAVSLLEQLKLFYEQQLFTDVLIVGTEPCHKAVLATCSSYFRAMFMSGLSESQOT 70
 Db 38 SNYAEKALKMTYMRSHGKLTIDVLEVKKELPKAKVVLASAPYFKAMFTGGLKESEMS 97
 QY 71 HVHLNVDPAATLQIIITAYTGNLAMDSTVEQLYETACFLQVEDVLORCREYLIIKKINA 130
 Db 98 RQDQGVCTANSRILYMYTQIRVTEYVQQLLPATNMFQVPNVIDACCAFLEKROLDP 157
 QY 131 ENCVRLLSFADLFSCBELKQSAKRMVHKFTAVYHODAFMQLSHDLIDILSSDNUNYEK 190
 Db 158 TNAIGIAHFAEQHGCVLEQKANKANFIERNFQVCOEEFLQLSAVQALIALIRDELAVOJE 217
 QY 191 EETVEAAMLMLENTSRSQYSSVLSQIRIDALSEVTQRAMFQGLPNDSSVVVVOGLY 250
 Db 218 EREVYNAVLKWKVYEDNRHCHMEHILGAVRC-----QPLTPN----- 255
 QY 251 KSMPEFKRLGTMTEEMWIFLEASSENPCSLYSV--CYSPOAEKYYKLCSPADLHKV 308
 Db 256 -----FLKEQM-----KNCDVLRVPAACREYLA-KIRF-----DLTLHKC 289
 QY 309 GTVV--TPDND--IYIAGG-----QVPLNKTNTNHSKTSKLOTAP-- 344
 Db 290 PGVKERTPTVTMIFVAGGFRRHSLDILEAVNVDDMTWTTLANLRI---PRSGLGAAPFK 346
 QY 345 -----RTVNCFY--WFD--AOONTMFPRTPLFVRIKPSLYNCEGCIYIAGID 388
 Db 347 GKPYAVGGRNNNIGSSYDSWDVDRYSATETWRPCAPMSVPHRRVGAVMDELMYAVG- 405

QY 389 SYGGEIENRTVERTEKDEWTVSPICAMQMSAAVVFDCIYVW-----TLNLMYC 441
 DB 406 SAGMEYH-NTVEYDPEDLDRMTLVOPMAKRLGVGVVVRLLYALGFGDNERLASVVC 464
 QY 442 YPRPSDWEMAMRORSRSPASAAFGDKIFYIGLHININSGIRLPSGTVDSS--YTV 499
 DB 465 YHPENNWSFLLPPLQGRSGAGVAAINQYIVVVG-----DETROLATV 509
 QY 500 EYDVNKNEMKMAANIPAKR 519
 DB 510 ERYDTENDTMDVAPIQIAR 529

RESULT 14

AAM47944
 ID AAM47944 standard; Protein; 575 AA.

AC AAM47944;

DT 01-MAR-2002 (first entry)

DE Human viscentric cycloctubulin protein 63.

KM Human; visicentric cycloctubulin protein 63; malignant tumour; haemopathy;
 KW development confusion disease; human immunodeficiency virus; HIV;
 KW infection; immune disease; inflammation.

OS Homo sapiens.

PN CN1311211-A.

PD 05-SEP-2001.

PF 02-MAR-2000; 2000CN-0111811.

PR 02-MAR-2000; 2000CN-0111811.

PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.

PI Mao Y, Xie Y;

DR WPI; 2002-049907/07.

DR N-PSDB; ABA05463.

PT New human cycloctubulin 63 polypeptide and encoding polynucleotide useful
 for treating tumour, haemopathy and human immunodeficiency virus -
 PS Claim 1, Page 27-28 (Disclosure); 35pp; Chinese.

CC The invention relates to human visicentric cycloctubulin protein 63, its
 CC recombinant production, antagonist, encoding polynucleotide and
 CC application. The polypeptide is useful for treating malignant tumour,
 CC haemopathy, development confusion disease, human immunodeficiency virus
 CC infection, immune disease and various inflammations.

CC Sequence 575 AA;

Query Match 15.9%; Score 523; DB 23; Length 575;

Best Local Similarity 25.0%; Pred. No. 1.7e-39; Indels 114; Gaps 15;
 Matches 151; Conservative 101; Mismatches 237;

QY 2 STQDERQITVEYAVSLLEQLFYEQQLFTDIYLVETGTEPCPKKAVLATGSSYFPAFM 61
 DB 10 SSSSEFYQVNHABQFRKMGESYLKQOQCDVILIGNKRIPAHRVLVSSVDYFAMFT 69
 QY 62 SGLSEKQTHVHLRVNDAATLQIITVAYTGNLANDSVEQLYETACFLQVEDVLRGR 121
 DB 70 SVVCEAKQREIKWEGIDPALMDLVQFAYTGCELEKEDTIENLLAAACLLQ.PQVVEVC 129
 QY 122 EYLTKINENCRVLSPADLPSCSELKQSAKMVHNK7YAVYHQAFLSHDLIDIL 181
 DB 130 HFLMKLHPNSCLGIRPADAGCIELMKVAHSYTMENINEYVRNDEPILLPAEELHKL 189

QY 182 SSDNINVEKETREAAMLMEYNTSSSQVLSVLSQIRI-----DALSEVTOBAMFG 236
 DB 190 ASDDVNVEDEETIFPALMMVXYDMQSRCDLSMLAFIRLPLPPQLADLENHALFKN 249
 QY 237 LEPNKSIVVQGLYKSMK-----FFKR----- 260
 DB 250 DLECKLILAMKHYLLPERRTLMQSPRTKPKSTVGLTVAVGMDNNKGAATTIERYDLR 309
 QY 261 -----LGMTKEBMMIFIEASENPSLYSVCSYSPQAKYKLCSPPAD 304
 DB 310 TNLMTQAGMMNRRLQPGVAVVDDKLVYIGRDGDKTANTVECVNPKKTIVL--PPMS 367
 QY 305 LHKVGTVP--DNDIYIAGGVPLNKTNTNKSITSKLTQAFRTVNCFTWPDAAQNTWPK 363
 DB 368 THRHGIGTVLEGPYAVVG-----HDGMSYLVNTVER-----MDPQSOQMTFV 410
 QY 364 TFMLEFVRILKPSLVCCGITYIAGDSVGGELNRRIVERYDTEKDEWTVSPICAMQMSA 423
 DB 411 ASMSIARSTVGVALNGKLYSVGRD--GSSCLSMSEYDPTNTMNCAPM-CRRGCV 467
 QY 424 AVVVDH-CIYVW-----TLNLMYCYPFRSDSWEMAMRQTSRSPASAAFGD 469
 DB 468 GVATCDGFLYAVGVGDAPASNHCSRLDYVERIDPKTDTWVAPLSMRAVGVCLGSD 527
 QY 470 KIFYIGLHIATNSGIRLPSGTVDGSSV--TVEYDVNKNEMKMAANIPAKRYSDPCVRA 527
 DB 528 RLAVVG-----YDQGYLTNTMESYDPTNEMTQMAALNIGR-AGACV-- 569
 QY 528 VVI 530
 DB 570 VVI 572.

RESULT 15

AAM40225
 ID AAM40225 standard; Protein; 748 AA.

AC AAM40225;

DT 22-OCT-2001 (first entry)

DE Human polypeptide SEQ ID NO 3370.

KM Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.

OS Homo sapiens.

PN WO200153312-A1.

PD 26-JUL-2001.

PF 26-DEC-2000; 2000WO-US34263.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AU, Yang Y, Zhang J;
 PI Zhao QA, Goodrich R, Dermanac RT;

DR WPI; 2001-442253/47.
DR N-PSDB; AA159381.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 5; SEQ ID NO 3370; 10078bp; English.

PS
XX The invention relates to human nucleic acids (AA157798-AA161369) and
CC the encoded polypeptides (AA38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX

Sequence 748 AA;

Query Match 15.9%; Score 523; DB 22; Length 748;

Best Local Similarity 25.0%; Pred. No. 2.6e-39;

Matches 151; Conservative 101; Mismatches 237; Indels 114; Gaps 15;

QY 2 STDERQINTYAVSLLEQDLFEYQQLFTDILVIVGTEFPCHKVLAATCSSYFRAMFM 61
DB 183 SSSEEFYQAVHABQFRKMESTYLKQOQLDVLIVGNRKIPARHLVSSVDYFAMFT 242
QY 62 SGSESQOTVHLRNVDAATLQIIITAYTGNLANNDSTYEOUYETRAFLQVEDVLORCR 121
DB 243 SDVCEAKQOEIKMEGIDPNALMDLVQFAVYGCLELKEDITENLAAACLOLPQVVEVCC 302
QY 122 EYLKKINAENCVRLLSFADLFSCSELKQSAKRWHEKFTAVVHODAFMQLSHDLIDIL 181
DB 303 HFLMKLHPNSCIGIRAFADAGCIELMKVASHYTMENIMEVIRNOEFLLPRAEELKLL 362
QY 182 SSDNLVKEKETVREAAMLMLEYNTSRSQYLSVLSQIRI-----DALSEVTORAMFOG 236
DB 363 ASDDVVVPDEETIFHALMMWVKYDMQSRCDLSMLLAFIRLPLPQILADLENHALFKV 422
QY 237 LPPNDKSVVQGLYKSPK-----FFPR----- 260
DB 423 DLSCQKILIEAMKYHLRRLTMOBPRTPRKSTVGTLYAVGMDNNKGATTIEKYDLR 482
QY 261 -----LGMTKEEMMTFIEASSENPCSLVSSVCYSPQAEKVYKLCSPPAD 304
DB 483 TNLMIQAGMANGRLQPGVAVIDDKLFVIGRGDLKTLNTVBCYNPKTKTWVL--PMS 540
QY 305 LHKVGTVPD-DNDIYIAGQVPLKNTKTNHSTKSLQTAFTVNCFYWFDAQOQNTWFPK 363
DB 541 THRHGLGVTLLEGPIYAVGG-----HDCMSYLTNVER-----WDPSQOQTFV 583
QY 364 TPMLFPIKPSLVCCBGYIYAIGDSVGEINRRTVERRYTEKDEWTMVSPLPCAMQMSA 423
DB 584 ASMSIARSTVGAALNGKLYSVGRD--GSSCLSSMEYDPTNTKNMCAFM-CKRRGV 640
QY 424 AVVVDH-CIYVM-----TLNLTVCYFPRSDSVEMAMQTSRSPASAAAFGP 469
DB 641 GVAITCGFLYAVGHDAPASNHCSRLLDYERIDPKTDITWYAPLSMPRDANGVCLLD 700
QY 470 KLFYIGLHATNSGIRLPSGTVDGSSV--TVEIYDVNKNEMKMAANIPAKRYSDPCVRA 527
DB 701 RLAVAVGG-----YDQTYLNTMESYDPQTNWETOMASLNIGR--AGACV--- 742
QY 528 VVI 530
|||

DB 743 VVI 745

Search completed: July 14, 2003, 18:14:54
Job time : 24.5998 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:10:22 ; Search time 13.7827 Seconds

(without alignments)
4345.419 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3292
Sequence: 1 MSTODERQINTSYAVSLLEQ.....STDGTEEFELDGEWVLPV 623

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1886	57.3	348	2 T17304	hypothetical prote
2	559	17.0	1476	2 A45773	hypothetical prote
3	497.5	15.1	531	2 T33099	hypothetical prote
4	438	13.3	618	2 T33222	kelch-like protein
5	415	12.6	836	2 T33322	hypothetical prote
6	373.5	11.3	625	2 T25373	hypothetical prote
7	327	9.9	564	2 C42523	A55R protein - vac
8	324	9.8	564	2 J01792	Salp17R protein -
9	312.5	9.5	559	2 S24675	p65 protein - Ect
10	292	8.9	585	2 T37216	caliclin - human (f
11	284	8.6	588	2 T46003	caliclin - bovine
12	244.5	7.4	509	1 WMVZMX	M9-R protein - myx
13	235.5	7.2	817	2 T24063	hypothetical prote
14	232	7.0	512	1 WZVZB6	59k HindIII-C proc
15	227	6.9	512	2 C42504	C2L protein - vacc
16	226	6.9	519	2 H88478	protein F47D12.7 l
17	216	6.6	480	2 G36213	F3L protein - vacc
18	209	6.3	476	2 T30781	kelch protein homo
19	208	6.3	480	2 H42506	F3L protein - vacc
20	203	6.2	251	2 C36418	M-TR protein - myx
21	185	5.6	552	2 T54388	L2TR-1 - human
22	183.5	5.6	518	2 S42387	MTP protein homol
23	182.5	5.5	581	2 T31745	hypothetical prote
24	178.5	5.4	274	2 S44616	C08C3.2 protein -
25	177.5	5.4	524	2 JC7090	germ cell-lesse pro
26	175.5	5.3	424	2 T39310	zinc finger protei
27	175.5	5.3	602	2 T21980	hypothetical prote
28	171.5	5.2	586	2 D84710	hypothetical prote
29	170.5	5.2	794	2 S59069	Z13 protein - mous

30	169.5	5.1	480	2 T20168	hypothetical prote
31	168	5.1	570	2 T24152	hypothetical prote
32	168	5.1	1466	2 T33791	hypothetical prote
33	166.5	5.1	410	2 T15442	hypothetical prote
34	166	5.0	320	2 T27372	hypothetical prote
35	166	5.0	484	2 T39200	zinc finger protei
36	162.5	4.9	410	2 S44626	C50C3.8 protein -
37	158	4.8	320	2 T31948	hypothetical prote
38	154	4.7	326	2 E86351	hypothetical prote
39	153.5	4.7	736	2 T00023	transcription fact
40	152.5	4.6	502	2 S16442	MTP protein - mou
41	152	4.6	551	2 T50663	RCCL-like G exchan
42	151	4.6	242	2 T31950	hypothetical prote
43	151	4.6	301	2 T33335	hypothetical prote
44	149.5	4.5	301	2 T32803	hypothetical prote
45	148	4.5	290	2 T32096	hypothetical prote

ALIGNMENTS

RESULT 1

T17304 hypothetical protein DKFP566C134.1 - human (fragment)

C/Specter: Homo sapiens (man)

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C/Accession: T17304

R: Bloembergen, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gasteiger, J.; Wilmann, S.

submitted to the Protein Sequence Database, September 1999

A/Reference number: Z18724

A/Accession: T17304

A/Status: preliminary

A/Molecule type: RNA

A/Residues: 1-348 <BIO>

A/Cross-references: EMBL:AL117562

A/Experimental source: fetal kidney; clone DKFP566C134

C/Genetics:

A/Note: DKFP566C134.1

Query Match 57.3%; Score 1886; DB 2; Length 348;

Best Local Similarity 99.4%; Pred. No. 2,4e-140;

Matches 346; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY	276	SENPCSLYSVCYSPQAEKRYKLCSPADLHKVGVTPDNDIYIAGGVPLNTKTNHS	335
DB	1	AENPCSLYSVCYSPQAEKRYKLCSPADLHKVGVTPDNDIYIAGGVPLNTKTNHS	60
QY	336	KTSKLQTAFTVNCVFPDAQNTWPKTPTMLFVRIKPSLVCCGGYIYVAGDSVGGELN	395
DB	61	KTSKLQTAFTVNCVFPDAQNTWPKTPTMLFVRIKPSLVCCGGYIYVAGDSVGGELN	120
QY	396	RTTVERVYDTEKEMTWSPCLPCAKWMSAAVVVHDCIYVNTLMLKCYFPFRSSWVMAMR	455
DB	121	RTTVERVYDTEKEMTWSPCLPCAKWMSAAVVVHDCIYVNTLMLKCYFPFRSSWVMAMR	180
QY	456	QTSRSPASAAAGDKIFVYGLHIAITNSGIRLPSGTVDSSTVETIYDNKMKEMKAANI	515
DB	181	QTSRSPASAAAGDKIFVYGLHIAITNSGIRLPSGTVDSSTVETIYDNKMKEMKAANI	240
QY	516	PAKRYSDPCVRAVVISNLSICVFMRETHLNERAKYVYTYQYDLELDKWSLRQHSERVLMDL	575
DB	241	PAKRYSDPCVRAVVISNLSICVFMRETHSNERAKYVYTYQYDLELDKWSLRQHSERVLMDL	300
QY	576	GSDFRCTVKGKLYPSCLESBPMKRPPTLYLSTDCTEBEELDGEWVLPV 623	
DB	301	GSDFRCTVKGKLYPSCLESBPMKRPPTLYLSTDCTEBEELDGEWVLPV 348	

RESULT 2

A45773 kelch protein, long form - fruit fly (Drosophila melanogaster)

N/Alternate names: ring canal protein kelch

N/Contains: kelch protein, short form

C:Species: Drosophila melanogaster
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 22-Oct-2001
C/Accession: A45773
R/Xue, F.; Cooley, L.
Cell 72, 681-693, 1993
A>Title: kelch encodes a component of intercellular bridges in Drosophila egg chambers.
A/Reference number: A45773; MUID:93201592; PMID:8453663
A/Accession: A45773
A/Molecule type: mRNA
A/Residues: 1-1476 <XUE>
A/Cross-references: GB:108483
A/Experimental source: embryo
A/Note: sequence extracted from NCBI backbone (NCBIN:127754, NCBI:P:127756)
C/Comment: This long form is produced by partial suppression of a TGA codon by an unknown C/Genic:
A/Gene: FlyBase:kel
A/Cross-references: FlyBase:FBgn001301
A/Superfamily: POZ domain homology
F:1-1476/Product: kelch protein, long form #status predicted <KPL>
F:1-688/Product: kelch protein, short form #status predicted <KPS>
F:142-244/Domain: POZ domain homology <POZ>
F:689/Region: opal stop codon

Query Match 17.0%; Score 559; DB 2; Length 1476;
Best Local Similarity 27.1%; Pred. No. 4e-35;
Matches 168; Conservative 104; Mismatches 250; Indels 98; Gaps 22;

QY 10 NTEYAVSLLEQLKLFYEQQLFTDVLIVGTEFPCHKWLATCSSYFRAMFMSGLSESKQ 69
DB 135 NEQHTARSPAMMBEMRKQKQLOLVDVLADVEITHAHMVLASCSFYFAFTS-FESRQ 193
QY 70 THVHLNVDATLQIIITVAYTGNLAMDSTVEQLYETACFLQVEDVLORCREYLKIKIN 129
DB 194 ARTLQSVARALBELIDYYTATVEVNEEDNVQLLTAANLLQLTVDRACCPFLQDQ 253
QY 130 AENCVRLLSPADLFSCBELKQSAKRVHEKFTAVTHODAFMQLSHDLIDILSSDNLNVE 189
DB 254 ASACLGIRFADPHACVETLNVAEITYEQHFNEVTOPEDEFLANSHQEVLSLINDRISIV 313
QY 190 KEETVEAMLMLENTESRSQYLSVLSQIRIDALSE--VTORAMQGLPRNDKSVVQ 247
DB 314 NEERVAECVIAIRYDPMKEOFTSLMEHVRLEPUSKEYTORV-----DKETLLE 365
QY 248 G-----LYKSMPEKFK-----PR--LGMTKEEMMIFIEASSENPCSLYSVCY 288
DB 366 GNIVCKNLLIEALTYHLPTETKSARTVPRKPGVMPR-----LLVIGGQPKAIRVEWY 421
QY 289 SPAEKYKLCSPPADLHKVTVTPNDIYIAGGOVPLKNTKTHSKTSKLTQATFTVN 348
DB 422 DLREEKYYQAAEMPNRCRSGSLVLDGK-VYAVG-----FNGS-----LRYTVQ 466
QY 349 CFYMPDAQONTWFPKTPMLFVRIRKPSLVCEGYIYAIGDSVGGELNRRTVERYDKE 408
DB 467 V---YDPATDQANCSMEARRSTLGVAAINGCIYAGG--PDGTGLSAAEMVDPRTDI 521
QY 409 WTVVSPPLPCAMQWAAVVDICIVM-----TLNLMYCYFPRSDSVEMAMQTSR 459
DB 522 WRIASGSTRSSVGVVGLIYAVGVGFTRCGLSSVERINPDITVNNVAEMSSR 561
QY 460 SFPSAAAFGDKITFYIGGLIATNSGIRLPSGTVDGSSV--TVBIYVNNKEMWMAANIPA 517
DB 582 SGAGVGVLLNLIYVVG-H-----DGPVRSVEAYCETSNMRSVADMSY 626
QY 518 KRYSDCVRAVVISNS--LCVPMRETHLNRAKYVYQVLEDRWSLRHISERVLMDL 575
DB 627 CRNN-----AGVAHGLLTVVGDDGTSLASVEVYCPD--SDSWRIIP-----ALMTI 674
QY 576 GRDPR--CTVGLKYPSCLEB 593
DB 675 GREYAGVCMIDK--PMXMEB 692

RESULT 3

T33099
hypothetical protein R12E2.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C/Accession: T33099
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 28-Jul-2000
R/Goela, D.; Scheet, P.
Submitted to the EMBL Data Library, May 1998
A/Description: The sequence of C. elegans cosmid R12E2.
A/Reference number: Z21281
A/Accession: T33099
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-531 <GOE>
A/Cross-references: EMBL:AF067219; PIDN:AACT17022.1; GSPDB:GN00019; CESP:R12E2.1
A/Experimental source: strain Bristol N2; clone R12E2
C/Genic:
A/Gene: CESP:R12E2.1
A/Map position: 1
A/Intons: 34/3; 60/3; 95/1; 186/3; 240/3
C/Superfamily: POZ domain homology
F:42-145/Domain: POZ domain homology <POZ>

Query Match 15.1%; Score 497.5; DB 2; Length 531;
Best Local Similarity 26.5%; Pred. No. 5.7e-31;
Matches 134; Conservative 82; Mismatches 226; Indels 63; Gaps 10;

QY 10 NTEYAVSLLEQLKLFYEQQLFTDVLIVGTEFPCHKWLATCSSYFRAMFMSGLSESKQ 69
DB 35 NSHFANSVLIQGLKRNVLCDVTLICGKRNNAHRYVLSCSFYLSMFTSQMAECIM 94
QY 70 THVHLNVDATLQIIITVAYTGNLAMDSTVEQLYETACFLQVEDVLORCREYLKIKIN 129
DB 95 REINMEIRPITLALIEFCYTGALIDDSNVQDILPAACLIQIHVEVQACCDYLKQKD 154
QY 130 AENCVRLLSPADLFSCBELKQSAKRVHEKFTAVTHODAFMQLSHDLIDILSSDNLNVE 189
DB 155 PNCIGIRAFADPHACVETLNVAEITYEQHFNEVTOPEDEFLANSHQEVLSLINDRISIV 214
QY 190 KEETVEAMLMLENTESRSQYLSVLSQIRIDALSE--ALSEVTORAMQGLPRNDK 243
DB 215 SEELVPSAVIQVNRHDSIKRTHLSMLSHVLPCTKPELVSVSEELVSDPASRL 274
QY 244 VVVOGLYKSM-----KFPKPRIGMTKEEMMIFIEASSENPCSLYSVCYSPAERY 295
DB 275 VDEAKVYLLPYERPMQGPRTKPRKPGSGSEVLA--CGWCSGDAIASI 323
QY 296 YKL-----GSPPADLHKVGT-VYTPNDIYIAGGOVPLKNTKTHSKTSKLTQAT 344
DB 324 ERLDPMKGGTTWKVAPMGKRCGVAVLENLLYAVG-----HDGQSYLNSIE 373
QY 345 RTVNCFYMPDAQONTWFPK-TPMLFVRIRKPSLVCEGYIYAIGDSVGGELNRRTVERYD 403
DB 374 R-----YDPTNOMSSVAPATCRISVGAARNGSIYAGGD--GESCLDVEKID 424
QY 404 TEKDEWTVVSPPLPCAMQWAAVVDICIVM-----TLNLMYCYFPRSDSVEMAMQ 456
DB 425 PKNEWAKYASMGTRBLGVSVNGCIYAVGSGNPAPLNTRVDRPVGKMEVRRPML 484
QY 457 TSRSFASAAAFGDKITFYIGGLIAT 481
DB 485 TRKRHLGTAIVDGYIYAVGGRDTT 509

RESULT 4

T37322
kelch-like protein kel-1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 28-Jul-2000
R/Ohmachi, M.
Submitted to the EMBL Data Library, February 1999
A/Description: Kel-1, a novel Kelch-related gene in C.elegans, is expressed in pharyngea
A/Reference number: Z21687

A/Accession: T3322
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-618 <OH>
 A/Cross-references: EMBL:AB023424; PIDN:BA082800.1
 R/Gajdary, S.
 submitted to the EMBL Data Library, March 1996
 A/Reference number: Z19209
 A/Accession: T20001
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-618 <WIL>
 A/Cross-references: EMBL:Z69902; PIDN:CAA93769.1; GSPDB:GN00020; CESP:C47D12.7
 A/Experimental source: clone C47D12
 C/Genetics:
 A/Gene: kel-1; C47D12.7
 A/Map position: 2
 A/Intons: 51/2, 97/2, 201/2, 255/2, 489/3, 538/3, 565/1, 599/3
 C/Superfamily: POZ domain homology
 F:42-160/Domain: POZ domain homology <POZ>

Query Match 13.3%; Score 438; DB 2; Length 618;
 Best Local Similarity 24.2%; Pred. No. 3.4e-26; Mismatches 206; Indels 114; Gaps 18;
 Matches 129; Conservative 83; Mismatches 206; Indels 114; Gaps 18;

```

QY 19 EQLKLFYEQLFTDYL---IVEGT-----EPPCHKMVLATCSYFRAMFMSG 63
DB 44 QNLNLMRCQDELCDVLEAVPIQGTSSBDGMDVQHGHNAHRVLSASSYFRAMFMSG 103
QY 64 LSESGQTHVHLNNDATLQIITAYTGNLAMDSTVEQLYETACFLOVEDYLQCRREY 123
DB 104 LRESGRIIPKEVVEVLSQLDIVWTGRMRIDEQNVQFILATSLQLTTCVRDCAFP 163
QY 124 LIKKINAENCVCALLSPADLFSCCELKQSAKAVEHEKTAHVYHQDAFMQSHDLIDISS 183
DB 164 MLELDMTVCVGNAEFARAHACHQLAAHQLYTRQHFVEIINEBELSLDKDAFCGLID 223
QY 184 DNLNKEETVEEAMLMLENTESRQYLSVLSQIRIDAL-SEVTOPEAFQGLPPNDK 242
DB 224 DRITVPEKXPVQAVLNVVAHEPSPRPPLAELMSVRLPLGDDVLPFKM-----RND 278
QY 243 -----SVVVOGLYSMPKPKPRCLMTEEMKIFLEASSENPCSLSYSCYSPQAE 293
DB 279 MIKSDAACINVIIEGMH-----QKRTKODI-----SSTEP-----HYPIQE 315
QY 294 ---KYKLCSP-PADLHKVGVTPNDIYIAGGVPLKTKTKNSKTSKLTQAPRTVC 349
DB 316 LNRKMFVAREPMBESCH-----IMVGGQAPKATVND----- 348
QY 350 FYWPDAAQQTWPKTPMLFVRIKPSLVCEGYIYAGDSVGGELNRRATVERVDEKDEM 409
DB 349 -LFPDQSLWSSCASLPRRCRSGVSMGNGVYTTG--FNGAQVVRVDYDPTDITW 404
QY 410 TWVSPPCAMQWASAAYVHDCIYV-----TLNLMYC-YF-PRSDSVEMAMKQTSRSPA 462
DB 405 RSANQMSARSTHGITTCCOVLVAVGFGDTGLASABYFDPHTGFWPLPMSSTRSSV 464
QY 463 SAAAGGDKIFYIGLHATNSGIRLPSGTVDGSSV---TVEIYDNKREMK 510
DB 465 GVAAFEEDLYAIGF-----DGVSKQCLNTVETPDRRAHKMR 501

```

RESULT 5

T3322
 hypothetical protein W029.2 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 28-Jul-2000
 C/Accession: T33222
 R/Rollfing, T.; Wohlmann, P.; Bauer, C.; O'Neal, D.
 submitted to the EMBL Data Library, May 1998
 A/Description: The sequence of C. elegans cosmid W029.
 A/Reference number: Z21304
 A/Accession: T33222

A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-836 <ROH>
 A/Cross-references: EMBL:AF067946; PIDN:AAC17684.1; GSPDB:GN00023; CESP:W029.2
 A/Experimental source: strain Bristol N2; clone W029
 C/Genetics:
 A/Gene: CESP:W029.2
 A/Map position: 5
 A/Intons: 87/3; 114/1; 144/1; 181/1; 312/2; 352/3; 441/3; 776/2
 C/Superfamily: POZ domain homology
 F:201-302/Domain: POZ domain homology <POZ>

Query Match 12.6%; Score 415; DB 2; Length 836;
 Best Local Similarity 22.4%; Pred. No. 3.4e-24; Mismatches 146; Conservative 88; Mismatches 223; Indels 196; Gaps 17;

```

QY 10 NTEVAVSLLEQLKLFYEQLFTDYLIEGTEFPCHKMVLATCSYFRAMFMSGSESG 69
DB 194 NQESKSKMEQWRILQTEELCDVELLVAGSVIRAHRYILAAASPYFRAMFTNGVEMKK 253
QY 70 THVHLRNVDAATLQIITAYTGNLAMDSTVEQLYETACFLOVED----- 115
DB 254 LTIELQDIPESVRIIVYITTDKIAITMNVHQLFTATVQLQVSDVCATHEILMKRY 313
QY 116 -----VLQRCREYLKIKINAENC 133
DB 314 DESRRQGFCEELARLQIKFPGFAGFLVPAPRPSLSPEMDVIVVACQFLATMTSHNC 373
QY 134 VRLSPADLFGCEELKQSAKAVEHEKTAHVYHQDAFMQSHDLIDISSDNLNKEET 193
DB 374 MSLVHPSDIYNTCLNLISSIEDFASQFRCIRKSPFNISFHHLSCLNRSDDLNVSEED 433
QY 194 VREAMLMV-----EYNTES---RSQYLSVLSQIRIDAL----- 225
DB 434 VFETIVQMLFETRLHLVGNMFLCEAVNSNIVKNSQCEKREIISMVLDAMTPSKRKE 493
QY 226 -----SEVTOPEAFQGLPPNDKSVVVOGLYSMPKPKPRCLMTEEMKIFLEASSEN 279
DB 494 SNHENTSEVSAAMCPSLTASSS-----STSPFKSVAG-----AIFCAGRGKA 539
QY 280 CSLYSV-CYSPQAEKVKYKLCSPPADLHKVGVTPNDIYIAGGVPLKTKTKNSKTS 338
DB 540 GGFSSVSAVYRNRQMIIEVDMQORRHVG-VVSNANLVAIG-----HDGTA 588
QY 339 KLQTFRTVNCVWDAQQTWPKTPMLFVRIKPSLVCEGYIYAGDSVGGELNRR 398
DB 589 HLAIEA-----PQSIQWKRIASMKTABRGIAVASIENVYAVGS--LDDTTCYKT 639
QY 399 VERVDEKDEWTVNPL-----PCAMQW-- 421
DB 640 VERVIDEEDWSTVADMVQGVGVAVIGRYLPAIGNDGTSLETCEPDPMDIKMR 699
QY 422 -----SAVVVHDCIYV-----TLNLMYC-YF-PRSDSVEMAMKQTSRSPA 464
DB 700 IASMKRRBAGSGVCLDYLYAIGFDDNAPLETCEPDPADKMTLIDKMSPPGVCV 759
QY 465 AAFGDKIYIGLHATNSGIRLPSGTVDGSSV---TVEIYDNKREMK 515
DB 760 AALGKRYAIGF-H-----DGSYLTNTVECYDPIANRKPAAEI 797

```

RESULT 6

T25373
 hypothetical protein T2789.4 - Caenorhabditis elegans
 C/Species: Caenorhabditis elegans
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 28-Jul-2000
 C/Accession: T25373
 R/Rollfing, T.; Wohlmann, P.; Bauer, C.; O'Neal, D.
 submitted to the EMBL Data Library, November 1996
 A/Reference number: Z20024
 A/Accession: T25373
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA

A:Residues: 1-625 <WIL>
A:Cross-references: EMBL:Z62059; PIDN:CAB04878.1; GSPDB:GN00021; CESP:T27E9.4
A:Experimental source: clone T27E9
C:GeneID: 3
A:Gene: CESP:T27E9.4
A:Map position: 3
A:Initiator: 36/2; 102/3; 132/1; 155/1; 214/3; 271/2; 305/3; 414/3; 483/1; 580/2
C:Superfamily: POZ domain homology
F:73-176/Domain: POZ domain homology <POZ>

Query Match 11.3%; Score 373.5; DB 2; Length 625;
Best Local Similarity 23.2%; Pred. No. 4e-21;
Matches 134; Conservative 97; Mismatches 239; Indels 107; Gaps 16;

```

QY 32 DIVLVEGTEPFGHKNVLAATCSYFRAMFSGSESQOTVHLRNDATLQITITAT 91
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 88 DVALLVENRKLSTNKHVLAATITPYFGMTLDMANKEINIEDNVEYVDLSPAT 147
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 92 GNLANMDSTVEQLYETACFLQVEDVLQRCREYLIKKNMENCVRLLSPADLFSC 150
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 148 GELRITTSNVQSIMLGANFQMLEVVOHCGNFLLRLHPNALSRBFCMVCVEE 207
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 151 SAGRMTEHKTATYHODAFWQLSHDLIDLSSDNINVEKEEYVREANMLLEYTES 210
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 208 MTDDYVQKHEMAVSXKEDFRLSLEDAIELRNDHLVDSSEQYVAAMWMLNCDV 266
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 211 QYLSVLSQIRIDALSEVTOGAWFQGLPNDKSVVVG--LYKSMKFFKPRLG 268
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 267 EQAKLILPCVRLDLS-----PFLSSIVASNPRIKDDIP--CRDLIDEAKDH 313
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 269 MFIENASSENPCSLYSSVCYSPQAEKVKYKLCSPADLHKVGVVTPNDIYIAG 328
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 314 LL-----PDRSLIKSFKCTP-----RLCQ-----IVEGLVAIVAG-- 344
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 329 NNTNHSKTSKQOTARTVACFWPDAQONTPEKTPMLEFVRIKPSLVCEGVIY 388
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 345 ---LMHOSKSSVEI-----YDPTSKKSPIDGMVTLRVRGAVANQROVAIG- 392
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 389 SVGGLNARTVERDYDEKEMTWSPPLPCAMQWSAAVVHDCIYV-----TL 441
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 393 -FNGQORLDLVEKFDVDTLKTTLSPLTRGRSALAAAFYNNRLYVCGGIDGL 451
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 442 YFPRSDSWEMANRQTSRSPASAAAFGDKIFYIGLH----- 478
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 452 YDINRWVMEAGPPEMNRSPAAGTVI--DKHLYICGHHDGQIPASVERLDT 510
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 479 -----IANSGRILPSGTVDSSV--TVEIYDANKENKMAANIPAKYSD 526
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 511 SMIQRCRCRGAATFKKIKYVAGYDTSFLKSVYVDPEKKWSPVSPMMNR-- 565
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 527 AVVISNSLCVFMRETHLNERAKVVTQYDELDRMSL 563
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 566 VSLVSTNEGIFAAGFDGENNLCSMEOYDVIDMSWIT 602
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 7

C42523
A:Residues: 1-564 <JOH>
A:Cross-references: A55R protein; POZ domain homology
F:7-110/Domain: POZ domain homology <POZ>
A:Status: preliminary
A:Molecule type: DNA
A:Reference number: A33172
A:Accession: C42523
A:Experimental source: strain WR
A:Title: Vaccinia virus homologues of the Shope fibroma virus inverted terminal repeat p
A:Reference number: A38550; MUID:91111982; PMID:1846491
A:Accession: C38550
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-564 <HOW>
A:Cross-references: GB:M58052
C:Superfamily: A55R protein; POZ domain homology
F:7-110/Domain: POZ domain homology <POZ>

Query Match 9.9%; Score 327; DB 2; Length 564;

Best Local Similarity 21.8%; Pred. No. 1.5e-17;
Matches 125; Conservative 90; Mismatches 221; Indels 138; Gaps 18;

```

QY 17 LLEQLKPEEQQLPFDIVIVIGTEPFGHKNVLAATCSYFRAMFSGSESQOTVHL 76
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 7 LIAVINGFPNRCRFGCINIVINDERIMAKLLTSGASEFSLFSNPFDSNVEYNASH 66
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 77 VDAATLQITITATYHODAFWQLSHDLIDLSSDNINVEKEEYVREANMLLEYTES 136
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 67 LDYQSVNDLIDYIGPLSTNDNVKYLSTADFIQIGALTECEYILKNCXNQCIDF 126
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 137 LSPADLFSCBELKQSAKRWTEHKTATYHODAFWQLSHDLIDLSSDNINVEKET 196
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 127 YIYADKYNKKKIESASFNTILQNLRLINDENFKLTIESMIKILSDMLNINKNE 186
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 197 AAMLMLEYTESQVLSVLSQIRIDALSEVTOGAWFQGLPNDKSVVVG--V 250
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 187 IILKMLESTQSCVTEL--LKCLNLSLS-----DQVTKSLYSHRLVSGIYEC 232
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 251 -----KSMKFFKPRLGMTKEEMMFIENASSENPCSLYSSVCYSPQAEKVK 299
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 233 ITFLNNIAFLDESFPYHSEL-----ISIGISNRKISINCYN----- 272
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 300 SPPADLHKVGT-----VTPNDIYIAG--QVPLK-----NNTNHSK 336
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 273 -----HKQWTEMTSSRRYRCSFAVAVLNIIYMGQDQSPYRSSKVIAVNTC 326
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 337 TSKLQTPARTVACFWPDAQO-----NTWFPKTP-----MLFVRI 371
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 327 YDIPDLKTPRSNCGGLADEYVYICGIRDDQSSLTSSIDKKPKSPYQKAKRREP 386
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 372 KPSLVCEGVIYIAGDSVSGELNARTVERDYDEKEMTWSPPLPCAMQWSAAVVH 430
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 387 DMGVAMNLNLIYVIGG--YKGDPTDALE--SLSDGWMKHQRLPIKMS--NM 442
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 431 IYV-----MTNLMTCYFPRSDSWEMANRQTSRSPASAAAFGDKIFYIGL 477
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 443 IYISGYNNSVYVNIYSLVSNPIYDEWTLSSLNIPRINPALMSAHNKL 502
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY 478 HIATNSGRILPSGTVDSSVTEIYDANKENKMAANIPAKYSD 511
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 503 DVRTN-----TSEYDKEKDCWTL 521
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

RESULT 8

JOI792
A:Residues: 1-564 <SMT>
A:Cross-references: DBJ:D11079; NID:g222717; PIDN:BA01828.1; PID:g222743
A:Experimental source: strain WR
A:Title: Vaccinia virus homologues of the Shope fibroma virus inverted terminal repeat p
A:Reference number: A38550; MUID:91259063; PMID:2045793
A:Accession: JOI792
A:Molecule type: DNA
A:Residues: 1-564 <SMT>
A:Cross-references: DBJ:D11079; NID:g222717; PIDN:BA01828.1; PID:g222743
A:Experimental source: strain WR
A:Title: Vaccinia virus homologues of the Shope fibroma virus inverted terminal repeat p
A:Reference number: A38550; MUID:91111982; PMID:1846491
A:Accession: C38550
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-564 <HOW>
A:Cross-references: GB:M58052
C:Superfamily: A55R protein; POZ domain homology
F:7-110/Domain: POZ domain homology <POZ>

Db 266 FTYYVNNVLYCLGTL-----DGAFTCDVIA-----YLLTINEXSLAMEGH 308
Qy 369 VAIKPSLVCCGGTYIAGDSVGGELNR--RTVERDYDEKDEWTWSPLPCAMQWGAAYV 426
Db 309 YRRMSACLVNGTIYVGVGD---EENRLIGSEYVQPMEEH----- 348
Qy 427 VHDCTY---VMTLNMV-----CYFP-----RSDSWEMANRQTSRSP 461
Db 349 -HDAPYLAQANVETATVCYRNEELWIVGGTVDLVHPFTISAVKLTDRMRMSMEPLPRSG 407
Qy 462 AAAAAAGDITFYIGLHIATNSGIRLPSCGVDSSTVTLIVD--NKNEKMAANIIPAG 519
Db 408 ATTIVYNNRLYCG-----GRHGGAYTNVYNYLDSRTMERVGDMAVNR 453
Qy 520 YSDPCRAVAVISNSLCVPMRETHL-----NERAKVYTYCYVDLEDRMSLRQHISERVLD 574
Db 454 RNPSC-----CVYKAIYVGGNTNAVEKY-----NGMKQGE----- 485
Qy 575 LGRDFRCVYKUL--YPSG 590
Db 486 -----VGNISTYPAC 495

RESULT 13

T24063
hypothetical protein R09A8.3 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C/Accession: T24063
R.Wilkinson, U.
submitted to the EMBL Data Library, November 1995
A/Reference number: Z19836
A/Accession: T24063
A/Status: preliminary; translated from GB/EMBL/DDHU
A/Molecule type: DNA
A/Residues: 1-817 <MIL>
A/Cross-references: EMBL:Z68009; PIDN:CAA92005.1; GSPDB:GN00028; CESP:R09A8.3
A/Experimental source: clone R09A8
A/Genetics:
A/Gene: CESP:R09A8.3
A/Map position: X
A/Intons: 31/3; 77/3; 133/1; 198/2; 240/1; 282/3; 330/3; 440/1; 745/2; 779/1

Query Match 7.2%; Score 235.5; DB 2; Length 817;
Best Local Similarity 16.9%; Pred. No. 4.2e-10;
Matches 127; Conservative 123; Mismatches 269; Indels 233; Gaps 28;

Qy 17 LLEQLKLFYEQQLFTDILVIVGTEPPCHKVLAATCGSSYFRAMF----- 60
Db 80 LLDRLNTFRRNRRLCDVLFVKEREIFVHKVLAASPALDMFNEENESSQDEKSPSS 139
Qy 61 -----MSGLS-----ESKQTHVHNRVDAATLQIIITYATGNL 94
Db 140 SPTNGHTNGTNGNEVAHENGAPVSAQPKQSTTFYEFESQDFFCFELVAFATYANL 199
Qy 95 ANNDSTVEQLYETACFLQVEDVLRQCREYLKIKINENCVRLLSFADLSCEBELKOSAKR 154
Db 200 EISRRVAAELKYKAAVLAQMTPIVIRACAAVLAENLQLSNCGIRQANFNHDTLMKRVQ 259
Qy 155 MEHAKTAAVYHQA-----FMQLSHDI 176
Db 260 FIVDNDSIYNDSEKFTQLFVIOVRIIVPAEDGKVTNNANNNGLAEMALFYQNNPHDR 319
Qy 177 L--L--L--L--SDNLNVEKEFTVREAAAMLLEYNTESNQ----- 211
Db 320 AEHSIDLLCTCKTHILYMEENHLDCIEMDRSSVGCDDI--QDYKSGKDRKQVAAKATA 378
Qy 212 ---YLSVLSQ-----IRI-----DALSEVTORAMFQGLPRNDK--SVVYQGLYKS 252
Db 379 QEPITINSVQHRVNGAVPRLNLSRLPMWASNESLESAGTDSDDQDTEARLISHTQ 438
Qy 253 MPFFFP-----RLGNTKEEMMITEASSENPCSLYS-----VCYSPQA 292

Db 439 AFOYVALVLYRRLCVLSLQITDNEELL--RTREPCPVDSQKALLSLRISCTGKQ 495
Qy 293 EKYYKLSPPADLHKVGTVPDNDIYAG-----QVPIKATKTNHSTSKLQ 341
Db 496 KPELCSAPRC---SIGASFL-NGKIPVCGGIDRECLRSVEEYVAVNGKRN--VAATK 549
Qy 342 TAFRTNCFYV-----FDAQONTWPKPTMLFVRIKPSLVCCGG 380
Db 550 ABRGFDDCTIHGGKFAVAGSNGNNDLKSCEIYDPAKVMAVPLPIATARCHNGCATIDN 609
Qy 381 YIYALG---DSVGEELNRRYERYDTEKDEWTWSPLP-----CAMQMSAAVVND 429
Db 610 YVYCIIGSGFDQKVLDCERLDTTKLGTBAEWEPAWSEHPRYQAGVCTWK--GLVVAAG 668
Qy 430 CIYVNTLNMVYCFPRSDSWEMANRQTSRSPASAAAGDKIFYIGLHIATNSGIRLPS 489
Db 669 CDRWTCMSVEAFEDKRTANAKQLPRLQARRGAAVAVRDITLYVIG--HDGTQS----- 721
Qy 490 GTVDGSSVTEIYDVNKNEMKMAANIIPAKRYSDPCRAVVISNSL--CV-----PMRE 540
Db 722 --LD---TVEILDSPSSQWRVGPPLTTPRAN---THAVVTAGNVIFCIGFNGVKFLDT 772
Qy 541 THLNERAKVITYQVLEDRMSLRQHISERV 572
Db 773 IELLEN-----EOIGWRNMQHCPQQL 794

RESULT 14

W21ZB6
59k HindIII-C protein - vaccinia virus (strain WR)
C/Species: vaccinia virus
C/Date: 31-Mar-1990 #sequence_revision 31-Mar-1990 #text_change 16-Jul-1999
C/Accession: F33348
R.Kotwal, G.J.; Moss, B.
Virology 167, 524-537, 1988
A/Title: Analysis of a large cluster of nonessential genes deleted from a vaccinia virus
A/Reference number: A94385; MUID:89073756; PMID:2849238
A/Accession: F33348
A/Molecule type: DNA
A/Residues: 1-512 <KOT>
A/Cross-references: GB:M2812; NID:g335691; PIDN:AAA69606.1; PID:g893315
A/Superfamily: vaccinia virus 59k HindIII-C protein

Query Match 7.0%; Score 232; DB 1; Length 512;
Best Local Similarity 21.5%; Pred. No. 3.9e-10;
Matches 116; Conservative 85; Mismatches 240; Indels 98; Gaps 22;

Qy 33 IVLIVGTEPPCHKVLAATCGSYFRAMFMSGLS--KQTHVHNRVDAATLQIIITYAY 90
Db 4 VIFSINGELIQVKEII-TASPY--NFKRIQDHLDKDEALILINGJNVHAFESLDYIR 59
Qy 91 TGNLAMDSTVQLYETACFLQVEDVLRQCREYLKIKINENCVRLLSFADLSCEBELKQ 150
Db 60 WKKINTINNVEMILVAIIIDVPPVVDLCVTMTIHNNINSTCIRMFNSKRYGKIKLYN 119
Qy 151 SAKRWVHKFTAVYHQAQFMQLSHDLIDILSSDNLNVEKEFTVREAAAMLLEYNTES-- 208
Db 120 AAMSEIINNITAVTSDPREKLSKDELTITISHEVNVNVNHEVTMLLKWHKQPNVD 179
Qy 209 -----RSQYLSVLSQIRIDALSEVTORAFQGLPRNDKSVVYQGL-YKSNPKFK 258
Db 180 IINILHPKMTNTMENAISLGL-TISKSTKPTVBNIGKH--IVYIKSDYISTITHYS 236
Qy 259 PR-----LGNTKEEMMITEASSENPCSLY-----SSVCYSQAQKVKYKLSG 301
Db 237 PRTEYWTIVGNDRQ--FYNNANVLNC-LYIIGGMINNRHAYSVSRVDLRTKKMTVTN 292
Qy 302 PADLHKVGTVPDNDIYAG--QVPLKATKTNHSHK--TSKLTQAFRTVNCFYWFDQON 358
Db 293 MSLSKSEVSTCNQDKLVYIGLSEFSGTGAELVKGHSK----- 333
Qy 359 TWPKPTMLFVRIKPSLVCCGGIYIAGDSVGGELNRRYERY--DTE--XDEWTM 411

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:15:27 ; Search time 14.6056 Seconds

(without alignments)
4965.877 Million cell updates/sec

Title: US-09-815-379-2

Sequence: 1 MSTDQRQINTEYAVSLLEQ.....STDGTEFELDEGVALPVPV 623

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Published Applications AA.*
2: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Prod. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3292	100.0	623	9	US-09-815-379-2 Sequence 2, Appl1
2	880	26.7	575	9	US-10-201-481-17 Sequence 17, Appl
3	565.5	17.2	609	9	US-10-097-065-237 Sequence 237, App
4	509	15.5	532	9	US-09-374-046A-78 Sequence 78, Appl
5	455	13.8	538	10	US-09-764-864-1250 Sequence 1250, App
6	374	11.4	287	10	US-09-864-761-42836 Sequence 42836, A
7	340.5	10.3	307	10	US-09-764-864-1222 Sequence 1222, Ap
8	336.5	10.2	337	9	US-09-925-300-1708 Sequence 1708, Ap
9	337	9.9	216	9	US-10-106-698-5100 Sequence 5100, Ap
10	233.5	7.1	224	10	US-09-764-864-1243 Sequence 1243, Ap
11	231	7.0	521	9	US-10-106-698-6227 Sequence 6227, Ap
12	228	6.9	156	10	US-09-925-300-1039 Sequence 1039, Ap
13	205.5	6.2	113	9	US-10-211-059-313 Sequence 313, App
14	202	6.1	379	9	US-10-106-698-5372 Sequence 5372, App
15	199	6.0	168	10	US-09-764-864-1223 Sequence 1223, Ap
16	199	6.0	385	10	US-09-925-302-624 Sequence 624, App
17	195.5	5.9	478	9	US-10-014-137-2 Sequence 2, Appl1
18	192.5	5.8	109	10	US-09-107-058-7 Sequence 7, Appl1
19	192.5	5.8	109	10	US-09-761-117-7 Sequence 7, Appl1

20	187	5.7	114	9	US-10-211-059-319 Sequence 319, App
21	180.5	5.5	115	9	US-10-211-059-322 Sequence 322, App
22	168.5	5.1	108	9	US-10-211-059-316 Sequence 316, App
23	168.5	5.1	111	9	US-10-211-059-321 Sequence 321, App
24	167	5.1	527	9	US-10-211-059-3 Sequence 3, Appl1
25	165	5.0	171	10	US-09-764-864-1225 Sequence 1225, Ap
26	163.5	5.0	111	9	US-10-211-059-314 Sequence 314, App
27	163.5	5.0	180	10	US-09-764-864-1224 Sequence 1224, Ap
28	163	5.0	123	10	US-09-867-550-828 Sequence 828, App
29	162.5	4.9	105	10	US-09-764-864-1237 Sequence 1237, Ap
30	153.5	4.7	195	10	US-09-764-864-1190 Sequence 1190, Ap
31	153	4.6	117	10	US-09-864-761-45806 Sequence 45806, A
32	151	4.6	117	9	US-10-097-065-664 Sequence 664, App
33	151	4.6	165	9	US-10-097-065-256 Sequence 256, App
34	148	4.5	104	9	US-10-164-359-2 Sequence 2, Appl1
35	147.5	4.5	1259	9	US-10-260-715-8 Sequence 8, Appl1
36	144	4.4	107	9	US-10-164-359-8 Sequence 316, App
37	143	4.3	114	9	US-10-211-059-318 Sequence 318, App
38	143	4.3	243	12	US-10-076-622-507 Sequence 507, App
39	143	4.3	243	12	US-10-007-805-507 Sequence 507, App
40	140	4.3	108	9	US-10-164-359-7 Sequence 7, Appl1
41	140	4.3	151	10	US-09-764-864-1231 Sequence 1231, Ap
42	138.5	4.2	641	9	US-10-108-605-61 Sequence 61, Appl1
43	138	4.2	158	10	US-09-764-864-1227 Sequence 1227, Ap
44	138	4.2	199	10	US-09-764-864-1228 Sequence 1228, Ap
45	136.5	4.1	354	10	US-09-731-872-416 Sequence 416, App

ALIGNMENTS

RESULT 1	
US-09-815-379-2	
Sequence 2, Application US/09815379	
Publication No. US20030073613A1	
GENERAL INFORMATION:	
APPLICANT: RASTELLI, LUCA	
APPLICANT: GERITSEN, MARY	
TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS	
TITLE OF INVENTION: ENCODING THE SAME	
FILE REFERENCE: 10716/35	
CURRENT APPLICATION NUMBER: US/09/815.379	
CURRENT FILING DATE: 2001-03-22	
PRIOR APPLICATION NUMBER: 60/191,134	
PRIOR FILING DATE: 2000-03-22	
NUMBER OF SEQ ID NOS: 17	
SOFTWARE: Patentin Ver. 2.1	
SEQ ID NO 2	
LENGTH: 623	
TYPE: PRT	
ORGANISM: Homo sapiens	
US-09-815-379-2	
Query Match	100.0%; Score 3292; DB 9; Length 623;
Best Local Similarity	100.0%; Prod. No. 1.86-284;
Matches 623; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MSTDQRQINTEYAVSLLEQLKLFYEQQLFDIVIVIGETSPCKRMVLATSSYFRANF 60
DB	1 MSTDQRQINTEYAVSLLEQLKLFYEQQLFDIVIVIGETSPCKRMVLATSSYFRANF 60
QY	1 MSTDQRQINTEYAVSLLEQLKLFYEQQLFDIVIVIGETSPCKRMVLATSSYFRANF 60
DB	1 MSTDQRQINTEYAVSLLEQLKLFYEQQLFDIVIVIGETSPCKRMVLATSSYFRANF 60
QY	61 MSLGESKQTHVHLNVDATLQIITTAAYGNLNMNSTVQLYETACFLQVEDLQRC 120
DB	61 MSLGESKQTHVHLNVDATLQIITTAAYGNLNMNSTVQLYETACFLQVEDLQRC 120
QY	121 REYLLIKINAENCVALSPADLFSCBBLKQSKRMVEHKFTAVYHQDAFMQISHDLIDI 180
DB	121 REYLLIKINAENCVALSPADLFSCBBLKQSKRMVEHKFTAVYHQDAFMQISHDLIDI 180
QY	181 LSSDLNVEKETTVEAMLMLEFYNTESRSQVLSVLSQIRDLASEVTQAPMFGGLPEN 240
DB	181 LSSDLNVEKETTVEAMLMLEFYNTESRSQVLSVLSQIRDLASEVTQAPMFGGLPEN 240

QY 241 DKSVMVGLYKSNPKFKPRLGNTKEEMMIFIEASSENPCSLYSVCYSPOAEKVKYKCS 300
DB 241 DKSVMVGLYKSNPKFKPRLGNTKEEMMIFIEASSENPCSLYSVCYSPOAEKVKYKCS 300
QY 301 PPADLHKVGVVTPNDNDIYIAGGVPLKNTKTNHSTKSLQTAFTVNCFCYWDAAQNTM 360
DB 301 PPADLHKVGVVTPNDNDIYIAGGVPLKNTKTNHSTKSLQTAFTVNCFCYWDAAQNTM 360
QY 361 PPKTPMLFVRIKPSLVCCEGYIYAIGDSVGGELNRTVRYDTEKDEMTWSPPLPCANO 420
DB 361 PPKTPMLFVRIKPSLVCCEGYIYAIGDSVGGELNRTVRYDTEKDEMTWSPPLPCANO 420
QY 421 WSAAVVHDCIYVMTNLAMYCFPRSDSWEMAMRQTSRSPASAAAGDKIFYIGLHIA 480
DB 421 WSAAVVHDCIYVMTNLAMYCFPRSDSWEMAMRQTSRSPASAAAGDKIFYIGLHIA 480
QY 481 TNSGIRLPGSTVGVSSVTVETIYVNNKEMMAANIPAKRSDPCRAVVISNLCVPMRE 540
DB 481 TNSGIRLPGSTVGVSSVTVETIYVNNKEMMAANIPAKRSDPCRAVVISNLCVPMRE 540
QY 541 THINERAKVYTYOYDELDRLMSLRHISERVLMDLGDFPCTVGLYPSCLSESPMKPPT 600
DB 541 THINERAKVYTYOYDELDRLMSLRHISERVLMDLGDFPCTVGLYPSCLSESPMKPPT 600
QY 601 YLFSTDGTBEFELDGEWVALPV 623
DB 601 YLFSTDGTBEFELDGEWVALPV 623

RESULT 2
US-10-201-481-17
Sequence 17, Application US/10201481
Publication No. US20030119024A1
GENERAL INFORMATION:
APPLICANT: Linsley, Peter
APPLICANT: Mao, Mao
APPLICANT: Biery, Matthew
TITLE OF INVENTION: Genes and Proteins Associated with T-Cell Activation
FILE REFERENCE: 9301-133-999
CURRENT APPLICATION NUMBER: US/10/201,481
CURRENT FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/306,968
PRIOR FILING DATE: 2001-07-20
NUMBER OF SEQ ID NOS: 19
SOFTWARE: PatentIn version 3.0
SEQ ID NO 17
LENGTH: 575
TYPE: PRT
ORGANISM: Homo sapiens
US-10-201-481-17

Query Match 26.7%; Score 880; DB 9; Length 575;
Best Local Similarity 33.3%; Pred. No. 1e-69;
Matches 198; Conservative 110; Mismatches 247; Indels 40; Gaps 10;

QY 13 YAVSLLEQLKLFYEQQLFTDIVIVE-GTEFPGHKWLATCSSYFRAMFNGSSEKQTH 71
DB 13 YAVSLLEQLKLFYEQQLFTDIVIVE-GTEFPGHKWLATCSSYFRAMFNGSSEKQTH 71
QY 5 HAGSLKQLKTMVDEGLTIDIVEVDHGTFSCHRNVLAMISPYFSMTSGLTSTQKE 64
DB 5 HAGSLKQLKTMVDEGLTIDIVEVDHGTFSCHRNVLAMISPYFSMTSGLTSTQKE 64
QY 72 VHLRNDATLQIITTYAYGNLAMDSTVEQLYETACPLQVEDVLTORCREYILKIKINE 131
DB 72 VHLRNDATLQIITTYAYGNLAMDSTVEQLYETACPLQVEDVLTORCREYILKIKINE 131
QY 65 VRIQVGEAESMDVLNAYATSRVILTEANVALFTAASIQIPIPSIDOCCKWIMISHDDP 124
DB 65 VRIQVGEAESMDVLNAYATSRVILTEANVALFTAASIQIPIPSIDOCCKWIMISHDDP 124
QY 132 NCYRLLSFADLFSCBELKOSAKRMVEHKTAAVYVHODAFMQLSHDLIDILSSONLWVEK 191
DB 132 NCYRLLSFADLFSCBELKOSAKRMVEHKTAAVYVHODAFMQLSHDLIDILSSONLWVEK 191
QY 125 NSIGVFLEADHYGHQELGDRSKYIRKFLCVTKQEFQDTQOLISLSDSDLVNDR 184
DB 125 NSIGVFLEADHYGHQELGDRSKYIRKFLCVTKQEFQDTQOLISLSDSDLVNDR 184
QY 192 ETVREAAMLLEVNTESRSQYLSVLSQ-TRIDALSEVTORAWFOGLPPNDKSVVQGLY 250
DB 192 ETVREAAMLLEVNTESRSQYLSVLSQ-TRIDALSEVTORAWFOGLPPNDKSVVQGLY 250
QY 185 EHVYESIIRMFEBHQNREHVLPEIFAKICRPLMBD---TIEKIPQFAQAIKSCV 240
DB 185 EHVYESIIRMFEBHQNREHVLPEIFAKICRPLMBD---TIEKIPQFAQAIKSCV 240
QY 251 KSNPKFFK---PLUGMTKEEMMIFIEASSENPCSLYSVCYSPOAEKVKYKCPDADLK 307
DB 251 KSNPKFFK---PLUGMTKEEMMIFIEASSENPCSLYSVCYSPOAEKVKYKCPDADLK 307

DB 241 EKGPSNTNGCTGRLGNTASEMILCFDAHAKHSKQKQVPCLDIITGRRVFKLCKPNDLRE 300
QY 308 VGTVVTPNDNDIYIAGGVPLKNTKTNHSTKSLQTAFTVNCFCYWDAAQNTMFPKTPML 367
DB 301 VGLVSPNDNDIYIAGGYP-----SSSEVSIDHQAENDFMVYDSTNMLSKPSLL 351
QY 368 FVRIKPSLVCCEGYIYAIGDSVGE--LNRRTVRYDTEKDEMTWSPPLPCANQMSAAV 425
DB 368 FVRIKPSLVCCEGYIYAIGDSVGE--LNRRTVRYDTEKDEMTWSPPLPCANQMSAAV 425
QY 352 RARIGCKLVYCCGKMYAIGGRVYEDGDRNLSKVSCEYDSRENCMTTVCAMPVAMEFHNAV 411
DB 426 VVHDCIYVMTNLAMYCFPRSDSWEMAMRQTSRSPASAAAGDKIFYIGLHIAITNSGI 485
QY 412 EYKKEKIYVQGEFFLEPEQKDYWGFLTPMTVPRIQGLAAVYKDSIYYIAG--TCGNQ 468
DB 486 RLPSGTVDSVTVETIYVNNKEMMAANIPAKRSDPCRAVVISNLCVPMRETHL-- 543
QY 469 RM-----FTVEAYDDELNKRTRKDPFCDQSNIPFLKVLVONKHLFVRATQVTV 519
DB 544 -----NERAKVYTYOYDELDRLMSLRHISERVLMDLGDFPCTVGLYPSCLSE 593
QY 520 EHVFRTRSKNSLVGYDDIADQMKVYETPDR-LMDLGRHFECAVAKLYPQCLQK 573

RESULT 3
US-10-097-065-237
Sequence 237, Application US/10097065
Publication No. US20030055236A1
GENERAL INFORMATION:
APPLICANT: Moore, Paul A. et al.
TITLE OF INVENTION: 110 Human Secreted Proteins
FILE REFERENCE: P2021P1
CURRENT APPLICATION NUMBER: US/10/097,065
CURRENT FILING DATE: 2002-03-14
PRIOR APPLICATION NUMBER: PCT/US98/27059
PRIOR FILING DATE: 1998-12-17
PRIOR APPLICATION NUMBER: 60/070,923
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/068,007
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/068,057
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/068,006
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/068,369
PRIOR FILING DATE: 1997-12-19
PRIOR APPLICATION NUMBER: 60/068,367
PRIOR FILING DATE: 1997-12-19
PRIOR APPLICATION NUMBER: 60/068,368
PRIOR FILING DATE: 1997-12-19
PRIOR APPLICATION NUMBER: 60/068,169
PRIOR FILING DATE: 1997-12-19
PRIOR APPLICATION NUMBER: 60/068,053
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/068,064
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/068,054
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/068,008
PRIOR FILING DATE: 1997-12-18
PRIOR APPLICATION NUMBER: 60/068,365
PRIOR FILING DATE: 1997-12-19
NUMBER OF SEQ ID NOS: 672
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 237
LENGTH: 609
TYPE: PRT
ORGANISM: Homo sapiens
US-10-097-065-237

Query Match 17.2%; Score 565.5; DB 9; Length 609;
Best Local Similarity 26.3%; Pred. No. 1.2e-41;
Matches 151; Conservative 109; Mismatches 222; Indels 93; Gaps 14;

Query Match 15.5%; **Score** 509; **DB** 9; **Length** 532

```

RESULT 5
US-09-764-864-1250
? Sequence 1250, Application US/09764864
? Patent No. US20020132753A1
? GENERAL INFORMATION:
? APPLICANT: Rosen et al.
? TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
? FILE REFERENCE: PT23
? CURRENT APPLICATION NUMBER: US/09/764,864
? CURRENT FILING DATE: 2001-01-17
? Prior application data removed - consult PALM or file wrapper
? NUMBER OF SEQ ID NOS: 1793
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO 1250
? LENGTH: 538
? TYPE: PRT
? ORGANISM: Homo sapiens
? FEATURES:
? NAME/KEY: SITE
? LOCATION: (263)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE
? LOCATION: (273)
? OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
? NAME/KEY: SITE

```



```

/ LOCATION: (332)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (346)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (347)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (351)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
/ NAME/KEY: SITE
/ LOCATION: (394)
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1250

Query Match      13.8%; Score 455; DB 10; Length 538;
Best Local Similarity 24.7%; Pred. No. 6.7e-32;
Matches 129; Conservative 97; Mismatches 206; Indels 90; Gaps 13;

QY 10 NTEVAVSLLEQLKLYEQQLFTDIVLVEGTEFPCHMVLATCSYPRAMFMSGLSESKQ 69
DB 84 NT-HAASILMSMSLRKSNLTLCVTLKVEQKDPARHIVLAACSDIFCAFTSELSKQ 142
QY 70 THVLRNVDAATLQIIITVAATGNLAMDSTVEQLYETACFLOVEDVLORCREYLKKIN 129
DB 143 PYVDIOQLTASTWEILLDFYETETVHTVENVOELLPACLLQKGVKQACCFLESQDL 202
QY 130 AENCVRLLSPADLFSCBELKQSAKRWVHKFTAVHODAMQJSHDILLIDLSNLYVE 189
DB 203 PSCNCLGRDPAETHNCVDLMQAAEVFSQKHPPEVVOHEBFILLISQGEVEKLKICDEIQVD 262
QY 190 KEETVEBAAMLMEYNTESRSOYLSSVLSQIRIDALSEVQGRAMFOGLPNDKSVVYQVL 249
DB 263 XEPPVEBAVAVXWVHAKKERESLPILQVYKMPILT----- 299
QY 250 YKSMPEFKPRKLCMTKEEMMIFTEASSENCPSLYSSVCSY--POAEKVYKLCSPPADLHK 307
DB 300 ----PRYITD-----VIDAEPFIRCSLQ---CRDLVDEAKKFF--LKPPELRSQM 339
QY 308 VGTVVTVPDNDI-YIAGQVVLKNTKTNHSTKSLQTRFRVNCFTVPDQONTWPFKTPM 366
DB 340 QGRTREXXLDKMYVSGG---FGDSRRHTSMER-----YDPNIDQMSMLGDM 382
QY 367 LFRIRKSLVCCGYIYAIGDSVGEINRTERYRDEKDEMTWSPPLCAMQWSAAVY 426
DB 383 QTRBEGGLVYVAGSVIYCLGGYDGLNLN--SVKIDPHGHTNTVTPMATTXSGAGVAL 440
QY 427 VHDCLVYM-----TLNLMYCYFPRSDSVWEMAMRQTSRSPASAAAFGDKIFYIGLHI 479
DB 441 LNDHIVVVGFPDGTALHLSVEAVNIRIDSWTVTSMTPPCVGTATLRLGLVALAG--- 497
QY 480 ATMSGIRLPSGTVDSSV--TVELIDVKNKEMKMANIPKR 519
DB 498 -----YDGNLSLISIECYDPIIDSWEVVSMGTQR 527

RESULT 6
US-09-864-761-42836
/ Sequence 42836, Application US/09864761
/ Patent No. US20020048763A1
/ GENERAL INFORMATION:
/ APPLICANT: Penn, Sharon G.
/ APPLICANT: Rank, David R.
/ APPLICANT: Hanzel, David K.
/ APPLICANT: Chen, Wenheng
/ TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
/ FILE REFERENCE: Aeonica-X-1
/ CURRENT APPLICATION NUMBER: US/09/864,761
/ PRIOR FILING DATE: 2001-05-23
/ PRIOR APPLICATION NUMBER: US 60/180,312
/ PRIOR FILING DATE: 2000-02-04

```

```

/ PRIOR APPLICATION NUMBER: US 60/207,456
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: US 09/632,366
/ PRIOR FILING DATE: 2000-06-03
/ PRIOR APPLICATION NUMBER: GB 24263.6
/ PRIOR FILING DATE: 2000-10-04
/ PRIOR APPLICATION NUMBER: US 60/236,359
/ PRIOR FILING DATE: 2000-09-27
/ PRIOR APPLICATION NUMBER: PCT/US01/00666
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00667
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00664
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00669
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00665
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00668
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00663
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00662
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00661
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: PCT/US01/00670
/ PRIOR FILING DATE: 2001-01-30
/ PRIOR APPLICATION NUMBER: US 60/234,687
/ PRIOR FILING DATE: 2000-09-21
/ PRIOR APPLICATION NUMBER: US 09/608,408
/ PRIOR FILING DATE: 2000-06-30
/ PRIOR APPLICATION NUMBER: US 09/774,203
/ PRIOR FILING DATE: 2001-01-29
/ NUMBER OF SEQ ID NOS: 49117
/ SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
/ SEQ ID NO 42836
/ LENGTH: 287
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ OTHER INFORMATION: MAP TO AL049870.1
/ OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.6
/ OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
/ OTHER INFORMATION: EXPRESSED IN HEL100, SIGNAL = 1.2
/ OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.96
/ OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.98
/ OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.98
/ OTHER INFORMATION: EST HUMAN HIT: U69560.1, EVALU 3.00e-55
/ OTHER INFORMATION: SWISSPROT HIT: Q04652, EVALU 3.00e-57
US-09-864-761-42836

Query Match      11.4%; Score 374; DB 10; Length 287;
Best Local Similarity 36.0%; Pred. No. 4.4e-25;
Matches 77; Conservative 48; Mismatches 89; Indels 0; Gaps 0;

QY 13 YAVSLLEQLKLYEQQLFTDIVLVEGTEFPCHMVLATCSYPRAMFMSGLSESKQTHV 72
DB 17 HSEQLQGLNLRRQHEHLCDDILRVGDVKIRAHKVVYLSVSYFFKAMFGNISEKENSEV 76
QY 73 HURNVDAATLQIIITVAATGNLAMDSTVEQLYETACFLOVEDVLORCREYLKKINEN 132
DB 77 BRQCIDETALQAIIVEAYATGVTFISQDYESLLPANNLLQITLVKECCAFLESQDPEIN 136
QY 133 CYRLISPADLFSCBELKQSAKRWVHKFTAVHODAMQJSHDILLIDLSNLYVEKE 192
DB 137 CGISRFATYTCRDLVLAATYICNPFVAVCQTEBFEFLTHADDEIVSNDCLNVAEE 196
QY 193 TVREAAMLMEYNTESRSOYLSSVLSQIRIDALS 226
DB 197 TVFYALBSWIKYDVOEROKITLQNLNSVRLPILS 230

```

```
RESULT 7
US-09-764-864-1222
; Sequence 1222, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT223
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1222
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (283)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (293)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1222

Query Match
Best Local Similarity 10.3%; Score 340.5; DB 10; Length 307;
Matches 66; Conservative 58; Mismatches 94; Indels 5; Gaps 1;

QY 32 DIVLIVEGTEPCHKMTATCGSYPRAMFMSGLSESKQTHVLRNVDAATLQIITTVAYT 91
DB 46 DIVLWQERKIPAHRVVLAASHFNLMTNMLSEKSEVELKDAEPDIIIEQLVEFAYT 105
QY 92 GNLANNDSTVEQLYACFLQVEDYLQRCREYLKIKINENCVRLLSPADLSCSELKQS 151
DB 106 ARISVNNXNVOSLLDAANOYQIEPVKMCVDLKEQVDSNCLGISVLAECLDCPELKAT 165
QY 152 AKRWVEHKFTAVYHODAFWQLSHDLLIDILSSDNLVKEETVREAAVMLLEYNTESRSQ 211
DB 166 ADDFIHQHFTVEYKTDDEFQLDVKRVTHLINDTLTVRAEDQYDAAVRMLKIDEPNRQP 225
QY 212 YLSSVLSQIRIDALSEVTORAMFQGLP-----PNDKSVVVOGL 249
DB 226 FVWDILAKVRPPLSKNPLSKTVQAEPLIQDNPECLKAVISGM 268

RESULT 8
US-09-925-300-1708
; Sequence 1708, Application US/09925300
; Patent No. US20020151681A1
; GENERAL INFORMATION:
; APPLICANT: Craig Rosen,
; APPLICANT: Steve Ruben
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA101
; CURRENT APPLICATION NUMBER: US/09/925,300
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05988
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1690
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1708
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (112)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
```

```
; LOCATION: (127)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (283)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-300-1708

Query Match
Best Local Similarity 10.2%; Score 336.5; DB 10; Length 337;
Matches 74; Conservative 62; Mismatches 113; Indels 31; Gaps 3;

QY 32 DIVLIVEGTEPCHKMTATCGSYPRAMFMSGLSKQTHVLRNVDAATLQIITTVAYT 91
DB 46 DIVLWQERKIPAHRVVLAASHFNLMTNMLSEKSEVELKDAEPDIIIEQLVEFAYT 105
QY 92 GNLANNDSTVEQLYACFLQVEDYLQRCREYLKIKINENCVRLLSPADLSCSELKQS 151
DB 106 ARISVNNXNVOSLLDAANOYQIEPVKMCVDLKEQVDSNCLGISVLAECLDCPELKAT 165
QY 152 AKRWVEHKFTAVYHODAFWQLSHDLLIDILSSDNLVKEETVREAAVMLLEYNTESRSQ 211
DB 166 ADDFIHQHFTVEYKTDDEFQLDVKRVTHLINDTLTVRAEDQYDAAVRMLKIDEPNRQP 225
QY 212 YLSSVLSQIRIDALSE-----VTORAMFQGLPNDKSVVVOGL 249
DB 226 FVWDILAKVRPPLSKNPLSKTVQAEPLIQDNPECLKAVISGMRYHLLSPEDREELVXGT 285
QY 250 YKSMKPKFRPKMGTKEMMIFIEASSENPCSLVSSVCS 289
DB 286 -----RPRKKHRYRIALF-GGSQPOSCRYFNPQDYS 316

RESULT 9
US-10-106-698-5100
; Sequence 5100, Application US/10106698
; Publication No. US20030109690A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
; FILE REFERENCE: PA005P1
; CURRENT APPLICATION NUMBER: US/10/106,698
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: PCT/US00/26524
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US 60/157,137
; PRIOR FILING DATE: 1999-09-29
; PRIOR APPLICATION NUMBER: US 60/163,280
; NUMBER OF SEQ ID NOS: 8564
; SOFTWARE: PatentIn Ver. 3.0
; SEQ ID NO 5100
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-106-698-5100

Query Match
Best Local Similarity 9.9%; Score 327; DB 9; Length 216;
Matches 63; Conservative 41; Mismatches 80; Indels 0; Gaps 0;

QY 9 INTEYAVSLLEQLKLFYEQOLFTDVLIVGTEPCHKMTATCGSYPRAMFMSGLSESK 68
DB 32 VSEHGQKILSVLQNFQGVYDPRKIMKEIILCHRCVLAAGSDFPRAMEVNMKEND 91
QY 69 OTHVLRNVDAATLQIITTVAYTGNLANNDSTVEQLYACFLQVEDYLQRCREYLKIKI 128
DB 92 DGSVITILSSKAVAFADYATGKTIYDQVEMFQLSFLOVSFLSKACSDPLIKSI 151
QY 129 NAENCVRLLSPADLSCSELKQSAKRWVEHKFTAVYHODAFWQLSHDLLIDILSSDNLV 188
DB 152 NLVNCLOLSTSDSYGSLFDPHALFVQHFSLFKSSDFLEMMFGVLOKCLBSDELAV 211
QY 189 EKKE 192
```

Db 212 PEEB 215

RESULT 10
US-09-764-864-1243
Sequence 1243, Application US/09764864

Patent No. US20020132753A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PT223

CURRENT APPLICATION NUMBER: US/09/764,864

Prior Filing Date: 2001-01-17

Prior application data removed - consult PALM or file wrapper

NUMBER OF SEQ ID NOS: 1792

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 1243

LENGTH: 224

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (15)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (94)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (185)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (212)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-764-864-1243

Query Match

Best Local Similarity 7.1%; Score 233.5; DB 10; Length 224;

Matches 55; Conservative 43; Mismatches 77; Indels 7; Gaps 4;

Db 10 NREYVSL--LELDKLFYEQLFTDVLIV--EGTE-FPCHKVVLACSSFFRAMFMSGLS 65

45 NTHSSVVLGGFDLRL--EGLLCDVTLVPGDDELFPVHRAMMASADVFCKMFTGGMK 101

66 ESKQTHVHLNRYDAATLQIITTYATGNLAMDSTVEQLYETPCFLOVEDVLCRCREYLI 125

102 EQLDMKIKLHGKVKVGLKTIIDITITAKLSLNDNDQDTLEAASFQILPVLDFCKVPLI 161

126 KKIINAEVRLISFADLFSCSELKQSAKRWVHKFTAVYHQDAFMQLSHDLIDILSSDN 185

162 SGVSLDNCVEVGRIANTVNLIEVKKYVNNFILKNFPALLSTGEFLKLPFXLAFLVLSNS 221

QY 186 LN 187

Db 222 LS 223

RESULT 11

US-10-106-698-6227

Sequence 6227, Application US/10106698

Publication No. US20030109690A1

GENERAL INFORMATION:

APPLICANT: Ruben et al.

TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide

FILE REFERENCE: PA005P1

CURRENT APPLICATION NUMBER: US/10/106,698

Prior Filing Date: 2002-03-27

Prior Application Number: PCT/US00/26524

Prior Filing Date: 2000-09-28

Prior Application Number: US 60/157,137

Prior Filing Date: 1999-09-29

Prior Application Number: US 60/163,280

Prior Filing Date: 1999-11-03

Query Match

Best Local Similarity 7.0%; Score 231; DB 9; Length 521;

Matches 101; Conservative 90; Mismatches 214; Indels 190; Gaps 21;

Db 122 EYLKIKINAEVRLISFADLFSCSELKQSAKRWVHKFTAVYHQDAFMQLSHDLIDL 181

2 DYLISRMVDTSISYRNFAFCWDSRLNKVDAYIQEHLLOISEEFLKLP--LKLQV 60

182 SSDNLNVEKEFTVREAMLMLEYNTERSQVLSVLSQIR-----IDALS 226

61 LEDNVCCLPSNGKLYRKVIVWVQRXIWNGDSLXXLMEEVQTLIYADHKLDGNLIDGQA 120

227 EV-----TORAWFOGLPNDKS----- 243

121 EVFGDDDDHICFVQKKPRENGHKQISSSTGCLSPNATVQSPKHEWKIVASEKTSNNT 180

244 ----VVQGLY-----KMPKFK-----PRGCM 263

181 YLCIAVLGIFCVIFLHGNPSQSPSTPTPKLSKLSFEMQDELIEKPMSPQYARSGL 240

264 TKEEM--MIFTEASENPSCLSYSSVCSPQAEKYLKCSF--PADLHKVGTVTDPND 317

241 GFAENKGLIAAGVYRRECC-LRTVECTNPHIDH--WFLAPKRTTRARQAVLA--GQ 295

318 IYIAGQVPLKNTKTNHSTKSLQTAFTVNGCFWFDQAQNTWPFKTPMLFVRIKPSLVC 377

296 LVVVG-----SNHSD-----DLSCGEYDSDIWDIMVPELRTRCNAGVCA 339

378 CEGYIYIAGSDVSGELNRRYERYDTEKDEMTWSPPLCAQMNAAVYVHDCIYV--- 434

340 INKGLYIYGSDPYQOKGLKNCDDVPVTKLMTSCAPINIRRHQSAVCELGGLYLTIGGA 399

435 ----TINLMYCYFPRSDSVEMAMRQTSRSPASAAFGDKPIFYIGLHATNSGIRLPSG 490

400 ESMNCLNRYVERINPENNITLIPNVARRGAVANGLKFLVCCGF----- 446

491 TVDGSV--TVEIYDVKNEWMAANIPAKRYSDPCVAVVISNLCVFMRETHLNERAK 548

447 --DGSIAISCVEMYDPTRENMKMGMTSPRSN--AGIATVGNIT----- 487

549 YTYOYDLEDRWSLRQHSERVLMDLGRDRCCTYGLKPLSCLSESPMKPPYTLF 603

488 YAVGGFD-----GNEFLNTV-EVYN--LESNEWSPYTKIF 519

RESULT 12

US-09-925-300-1039

Sequence 1039, Application US/09925300

Patent No. US20020151681A1

GENERAL INFORMATION:

APPLICANT: Craig Ruben

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA101

CURRENT APPLICATION NUMBER: US/09/925,300

```
/ CURRENT FILING DATE: 2001-08-10
/ PRIOR APPLICATION NUMBER: PCT/US00/05988
/ PRIOR FILING DATE: 2000-03-08
/ PRIOR APPLICATION NUMBER: 60/124,270
/ PRIOR FILING DATE: 1998-03-12
/ NUMBER OF SEQ ID NOS: 1890
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 1039
/ LENGTH: 156
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-925-300-1039

Query Match
Best Local Similarity 35.0%; Score 228; DB 10; Length 156;
Matches 43; Conservative 27; Mismatches 53; Indels 0; Gaps 0;

Qy 9 INTERAVALLEOQLFYEQQLFTDVLIVYGTBEPCHKQVLAACSSYFFAMNSGISESK 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 32 VSEDHGOQLSYLQNFREQNVFDFKIKMDEIIPCHRCVLAACSDFFRAMEFVNKKEKD 91

Qy 69 QTHVLRVNDATLQIIITVATGNLAMDSTVEQLYACFLOVEDVLORCREYLKIKI 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 92 DGSVITTLSSKAVKAFDPAITGKTKITDQNVEMFFQLSSFLQVSLSKACSDPLIKSI 151

Qy 129 NAE 131
   | |
Db 152 NLE 154

RESULT 13
US-10-211-059-313
/ Sequence 313, Application US/10211059
/ Publication No. US20030100495A1
/ GENERAL INFORMATION:
/ APPLICANT: Zhang, Jian
/ TITLE OF INVENTION: HUMAN NAC-1 PROTEIN
/ FILE REFERENCE: PB0149
/ CURRENT APPLICATION NUMBER: US/10/211,059
/ CURRENT FILING DATE: 2002-08-02
/ PRIOR APPLICATION NUMBER: US 60/311,034
/ PRIOR FILING DATE: 2001-08-08
/ NUMBER OF SEQ ID NOS: 322
/ SOFTWARE: Aecomica Sequence Listing Engine
/ SEQ ID NO 313
/ LENGTH: 113
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ OTHER INFORMATION: Consensus sequence of BTB motif
US-10-211-059-313

Query Match
Best Local Similarity 38.1%; Score 205.5; DB 9; Length 113;
Matches 43; Conservative 20; Mismatches 49; Indels 1; Gaps 1;

Qy 16 SLLEOLKLFYEQLFTDVLIVYGTBEPCHKQVLAACSSYFFAMNSGISESKQTHVLR 75
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 2 SLQNLINERNGEFCDVTLVVGKFKKAKAVLACSPFKLFFGNFKES-SSITTD 60

Qy 76 NYDAATLQIIITVATGNLAMDSTVEQLYACFLOVEDVLORCREYLKIKI 128
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 DVSVEFALBERITGKLEITENVELELADKQIGSLVDLGEFLIKNI 113

RESULT 14
US-10-106-698-5372
/ Sequence 5372, Application US/10106698
/ Publication No. US20030109690A1
/ GENERAL INFORMATION:
/ APPLICANT: Ruben et al.
/ TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide
/ FILE REFERENCE: PA005PI
```

```
/ CURRENT APPLICATION NUMBER: US/10/106,698
/ CURRENT FILING DATE: 2002-03-27
/ PRIOR APPLICATION NUMBER: PCT/US00/26524
/ PRIOR FILING DATE: 2000-09-28
/ PRIOR APPLICATION NUMBER: US 60/157,137
/ PRIOR FILING DATE: 1999-09-29
/ PRIOR APPLICATION NUMBER: US 60/163,280
/ PRIOR FILING DATE: 1999-11-03
/ NUMBER OF SEQ ID NOS: 8564
/ SOFTWARE: Patent In Ver. 3.0
/ SEQ ID NO 5372
/ LENGTH: 379
/ TYPE: PRT
/ ORGANISM: Homo sapiens
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (10)_
/ OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-106-698-5372

Query Match
Best Local Similarity 23.9%; Score 202; DB 9; Length 379;
Matches 92; Conservative 62; Mismatches 171; Indels 60; Gaps 15;

Qy 201 WLEYNTERSOYLSSVLSQIRIDALSB--VTQRAWFGGLPPND--KSVVQGL-YKSNP 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 13 WIKYDVEROKYLAQLNSVRLPLSLVKFLFRLYBANHLIRDRCKHLNBLAKYHFW 72

Qy 255 K-----FFKPLGKTKEMMIFEASSENPCSLYSVCSQAERKVTGLCSPPADL 305
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 EHRLSHQVLMTRPCA---PKVLCAGVGKSGFLFACLSVEMYFPQNSWIGLABLNTPR 129

Qy 306 HKVGTVPNDNDIYAGGOVPLKNTKTNHSTKSKQTAFRTVNCYWFDPDAQNTWPKPTP 365
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 130 YEGICVL-DQKVTYIG-----IATVPRGVITRKENSVEG--KNPDNTVTSLEK 179

Qy 366 MLFVRIKSLVCEGYIYAIGDSVSGELNRTTVERDYDEKDEWTVSPLPCAMQMSAAV 425
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 180 MNESSTLGVVVLAGELVALGG--YDQSYLQSVKVIPIKIKMQPVAPMTTRSCFALA 237

Qy 426 VVHDCIYVW-----TILMLCYFPRSDSVEMAKRQSRASAAARODKIFYGGLHI 479
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 VLDGMITVIAIGYGPAMNNSVERYPDSKDSWEVAMADKRIFHFGVNLGFIYVGG-- 294

Qy 480 ATNSGIRLPSTGVDSSTVVEIYDVNKNEMKMAANIPAKRYSDPCRAVVISNSLCV--- 536
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 --HNGVSHLS-----SIERYDPHONQWYCRMKKEPRIG--VGAADVINYLYVVG 341

Qy 537 FMRETHLNERAKYVTQYDLLEDRW 561
   : : : : :
Db 342 HSGSSYLN-----TVQKYDPISDTW 361

RESULT 15
US-09-764-864-1223
/ Sequence 1223, Application US/09764864
/ Patent No. US20020132753A1
/ GENERAL INFORMATION:
/ APPLICANT: Rosen et al.
/ TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
/ FILE REFERENCE: PT223
/ CURRENT APPLICATION NUMBER: US/09/764,864
/ CURRENT FILING DATE: 2001-01-17
/ Prior application data removed - consult PAM or file wrapper
/ NUMBER OF SEQ ID NOS: 1792
/ SOFTWARE: Patent In Ver. 2.0
/ SEQ ID NO 1223
/ LENGTH: 168
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-09-764-864-1223

Query Match
Best Local Similarity 6.0%; Score 199; DB 10; Length 168;
```

Best Local Similarity 27.8%; Pred. No. 7.8e-10;
Matches 42; Conservative 39; Mismatches 52; Indels 18; Gaps 2;

QY 11 TEAVASLEQLK-----FYEQLFTDIVLIVEGTEFPCHKVLATCSSYFRAMF 60
Db 16 TEKKLAAREAKLLAGFMGMNNRKKKTLCDVILWQERKIPAHRVLLAASHFFNLMP 75
QY 61 MSGLSSESKQTHVHLRVNDAATLQIITVAYTGNLANNDSTVEQLYETACFLQVEDVLRQ 120
Db 76 TTNMLSSKSPFEVELKDAEPDIIEQLVEFAVTARISVNSNNVQSLDANQYQIEPVKWC 135
QY 121 REYLIIKKINAEKNCVRLISPADLFSCEELKOS 151
Db 136 VDPLKEOVDAKNC-----LGEAEKVDOS 158

Search completed: July 14, 2003, 18:26:55
Job time : 16.6056 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:46 ; Search time 6.47994 Seconds
(without alignments)
3987.653 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3292
Sequence: 1 MSTDQERQINTEYAVSLLEQ.....STDGIEPELDGEMVALPPV 623

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt 40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	572	17.4	606	1 KRPI_RAT	Q9e330 ratu
2	569	17.3	1477	1 KRLC_DROME	Q04652 drosophila
3	565.5	17.2	609	1 KRLP_HUMAN	Q9y4m5 homo sapien
4	554	16.8	606	1 KRLP_HUMAN	Q06662 homo sapien
5	545.5	16.6	587	1 KHL2_HUMAN	Q9uh77 homo sapien
6	534	16.2	593	1 KHL2_HUMAN	Q95198 homo sapien
7	523	15.9	748	1 KHL1_HUMAN	Q9n164 homo sapien
8	516.5	15.7	589	1 ENCI_MOUSE	Q35709 mus musculu
9	516	15.7	751	1 KHL1_MOUSE	Q91714 mus musculu
10	511.5	15.5	589	1 ENCI_HUMAN	Q14682 homo sapien
11	505.5	15.4	718	1 KHL4_HUMAN	Q9c0h6 homo sapien
12	501	15.2	755	1 KHL5_HUMAN	Q96997 homo sapien
13	492	14.9	624	1 KEAP_HUMAN	Q14145 homo sapien
14	474	14.4	632	1 YDS4_HUMAN	Q9p2j3 homo sapien
15	473	14.4	624	1 KEAP_MOUSE	Q9z2x8 mus musculu
16	467	14.2	624	1 KEAP_RAT	P57790 ratu
17	455.5	13.8	584	1 IPP_HUMAN	Q9y573 homo sapien
18	455	13.8	639	1 YD05_HUMAN	Q9p2n7 homo sapien
19	445	13.5	652	1 YD84_HUMAN	Q9p293 homo sapien
20	443.5	13.5	539	1 Y466_HUMAN	Q9u1p4 homo sapien
21	418.5	12.7	584	1 IIP_MOUSE	P28515 mus musculu
22	399	12.1	597	1 GAN_HUMAN	Q9h2c0 homo sapien
23	345.5	10.5	465	1 Y795_HUMAN	Q94889 homo sapien
24	327	9.9	564	1 VAS5_VACCC	P21073 vaccinia vl
25	324	9.8	564	1 VAS5_VACCV	P24768 vaccinia vl
26	309.5	9.4	530	1 VC04_SPVKA	P32228 swinepox vl
27	287	8.7	588	1 CALI_HUMAN	Q13939 homo sapien
28	284	8.6	588	1 CALI_BOVIN	Q28068 bos taurus
29	273	8.3	623	1 Y711_HUMAN	P22611 myxoma viru
30	251	7.6	623	1 Y711_HUMAN	Q94819 homo sapien
31	244.5	7.4	509	1 VMT9_MYXVU	P08073 myxoma viru
32	234.5	7.1	500	1 VC13_SPVKA	P32206 swinepox vl
33	232	7.0	512	1 VC02_VACCV	P17371 vaccinia vl

34	227	6.9	512	1 VC02_VACCC	P21037 vaccinia vl
35	226	6.9	579	1 YR47_CAEBL	Q09563 caenorthabd1
36	216	6.6	480	1 VF03_VACCV	P24357 vaccinia vl
37	208	6.3	480	1 VF03_VACCC	P21013 vaccinia vl
38	202.5	6.2	441	1 YD40_HUMAN	Q9p203 homo sapien
39	199	6.0	374	1 SPOF_HUMAN	C43791 homo sapien
40	195	5.9	918	1 SCRA_TIMPO	Q25390 timulus pol
41	190.5	5.8	522	1 BTE3_HUMAN	Q9y2f9 homo sapien
42	188.5	5.7	532	1 BTE3_MOUSE	F58545 mus musculu
43	183.5	5.6	518	1 YNV6_CAEBL	P34569 caenorthabd1
44	180.5	5.5	525	1 YKV2_HUMAN	C9p270 homo sapien
45	178.5	5.4	274	1 YKV2_CAEBL	P34324 caenorthabd1

ALIGNMENTS

RESULT 1	ID	Query Match	Score	DB 1	Length
KRPI_RAT	1	17.4%	572	1	606
AC Q9ER30;	16-OCT-2001 (Rel. 40, Last sequence update)				
DT 16-OCT-2001 (Rel. 40, Last sequence update)					
DE Kelch-related protein 1 (Kel-like protein 23) (Sarcosin).					
GN KRPI.					
OS Rattus norvegicus (Rat).					
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.					
OX NCBI_TaxID=10116;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC TISSUE=Embryonic;					
RX MEDLINE=20180502; PubMed=10713668;					
RA Spence H.J., Johnston I.P., Bwart K., Buchanan S.J., Fitzgerald U.,					
RA Ozanne B.W.;					
RT "Krip1, a novel kelch related protein that is involved in pseudopod					
RT elongation in transformed cells.";					
RL Oncogene 19:1266-1276(2000).					
CC -1- FUNCTION: REQUIRED FOR PSEUDOPOD ELONGATION IN TRANSFORMED CELLS.					
CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY CYTOPLASMIC BUT CAN CO-					
CC LOCALIZE WITH F-ACTIN AT THE MEMBRANE RUFFLE-LIKE STRUCTURES AT					
CC THE TIPS OF TRANSFORMATION-SPECIFIC PSEUDOPODIA.					
CC -1- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN SKELETAL MUSCLE. ALSO					
CC FOUND IN HEART AND LUNG.					
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.					
CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.					
CC This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC the European Bioinformatics Institute. There are no restrictions on its					
CC use by non-profit institutions as long as its content is in no way					
CC modified and this statement is not removed. Usage by and for commercial					
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC or send an email to license@isb-sib.ch).					
CC -----					
DR EMBL: AJ293948; CAC08185.1; --					
DR InterPro: IPR000210; BTB_POZ.					
DR InterPro: IPR001798; Kelch.					
DR Pfam: PF00651; BTB; 1.					
DR Pfam: PF01344; Kelch; 5.					
DR SMART: SM00225; BTB; 1.					
DR PROSITE: PSS0097; BTB; 1.					
DR Cytokeleton; Repeat.					
KW DOMAIN	33	100			
FT REPEAT	346	398			
FT REPEAT	399	447			
FT REPEAT	448	495			
FT REPEAT	497	542			
FT REPEAT	544	599			
FT SEQUENCE	606 AA;	68213 MW;	08DB54E9298DACE CRC64;		

Beat Local Similarity 27.0%, Pred. No. 2-2e-34:		
Matches	157, Conservative	95, Mismatches 236, Indels 94, Gaps 13
QY	6 EROINTE--YAVSLLEO-LKLFYEOQLTDIVLIVEGTEPFCHKMVLTATGSSYFRAMFM	61
DB	4 ORELAEEELRLYOSTLLDQGLKDLBEKKRTDCTLWAGDKSPFCHRLITISACSPFREYFL	63
QY	62 SGLSEKQTHVHLRVNDAATLQIIITVATGMLAMNDSTVEDLOETACFLQVEDLQSCR	121
DB	64 SEIBEKKKEMALDNVDPALIDLIIIKYLSASIDLNDGVNDIPLASSNFQISVFTVCV	123
QY	122 EYLKIKINAEVCLLSPADLFCSEBLKOSAKRMVBHKTAVYHODAFMQLSHLLIDL	181
DB	124 SYLQRLPLGNCIALIRGLDLDCRLLISAEPFSDREVQCKEDEFMQLSPQELISVI	183
QY	182 SSDNLNVEKEETVREAMLMLEYNTESNOYSSVLSQRLDALSE-----	227
DB	184 SNDSLNVKEEVEFPAVMKMYTVDENRANLSEVPDCIRFLMBEKKFKDHEVEDIILK	243
QY	228 -----YTQAMFOGLPPNDKSVVVGUYK-----SNPKFEK--PRLGMTKE	266
DB	244 SNPEVQKKIKIVLKDAFAGLPEPSSAEKGGTGVNGDVEDLLPGYLANDIPRHGMFVK	303
QY	267 EMMIIEIASSENPCGLSSVCSPOAEKKYKLCSPPADLHKGTAVTPNDIYINGGVV	326
DB	304 DILFLVNDP-----AAVAYDPEMNECYLTALAEQIPRNHSSIVTQONQYVVGGLY-	354
QY	337 LKNYTNHSTSKLQTAFTVVCFWFPAQOQNTWPKTPLLFRIRKPSLVCEBG-	381
DB	355 -----VDEENKQDPLQ-----SYFQLDVNVSEWGLPPL-----PSARCFLGGEYVD	398
QY	382 -IYAAGDSVGEELMRRTYERDYDEKEDWTWSPLPCAMQMSAAVNHDCIYVM-----	434
DB	399 KIVYVAGDLOTEASLDVLCYDPVAAKMSEVKNLPIKYGHNVISHNGMIYCLGAGKTD	458
QY	435 --TLNLMCYPRSSWEMAMRQTSRSEFAAAAGGDKIFYIGGLHIAITNSGIRLPGSTV	492
DB	459 KKCTNRVFIYNPKGDMDLAPMKTPRSMFGVALIHKGKIVIAGV-----TE	505
QY	493 DGSSTVTEIYDVNKQEMKMAANIIPAKRISDPCVRVAVVINSGL	534
DB	506 DGLSASVAFDLKTKMKVEWMTPEFPERSS--ISLVSLAGSL	544

	RESULT 2	
KELC_2	DROME	STANDARD; PRT; 1477 AA.
ID	KELC_DROME	
AC	Q04652; Q04653; Q9VGA2;	
DT	01-FEB-1994 (Rel. 28, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	15-JUN-2002 (Rel. 41, Last annotation update)	
DE	Ring canal kelch protein [Contains: Kelch short protein].	
GN	KEL OR CG7210.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Mandibulate; Pancrustacea; Hexapoda;	
OC	Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;	
CC	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Embryo;	
RX	MEDLINE=93201592; PubMed=8453663;	
RA	Xue F., Cooley L.;	
RT	"Kelch encodes a component of intercellular bridges in Drosophila eggs	
RT	chambers.";	
RL	Cell 72:681-693(1993).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Berkeley;	
RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	

RA Sution G.G., Mortman J. R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazel R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Heltz R.G., Nelson C.R., Miklos G.L.G.,
RA Abail J.F., Agbayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Baen A.A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottler P.,
RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlike C., Davenport L.B., Davies P.,
RA de Pablos B., Detcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doug L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Dudson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flocker C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman I.J., Hernandez J.R., Houck J.,
RA Hostein D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalustin F., Karen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulip D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McInerason D.,
RA Mekulov G., Mleishna N.V., Nobary C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskeken D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spiter B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissensbach J.,
RA Williams S.M., Woodde T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RA Science 287:2185-2195(2000).
RL [3]
RN CHARACTERIZATION.
RP TISSUE=EMBYO;
RC MEDLINE=97236487; PubMed=9118811;
RX Robinson D.N., Cooley L.;
RA "Examination of the function of two kelch proteins generated by stop
RT codon suppression.";
RL Development 124:1405-1417(1997).
RL -1- FUNCTION: COMPONENT OF RING CANALS THAT REGULATES THE FLOW OF
CC CYTOPLASM FLOW FROM NURSE CELLS TO THE OOCYTE DURING OOGENESIS.
CC CYTOPLASM FLOW FROM NURSE CELLS TO THE OOCYTE DURING OOGENESIS.
CC BINDS ACTIN.
CC -1- SUBCELLULAR LOCATION: INNER SURFACE OF CYTOPLASMIC BRIDGES OR RING
CC CANALS PRESENT IN EGG CHAMBERS. SUBCORTICALLY IN IMAGINAL DISK
CC EPITHELIUM.
CC -1- TISSUE SPECIFICITY: BOTH PROTEINS ARE EXPRESSED IN OVARIES, MALE
CC TESTIS, OVACECTOMIZED FEMALES, CUTICLE, SALIVARY GLAND AND
CC IMAGINAL DISKS. KELCH ORF1 IS THE PREDOMINANT PROTEIN AND IS ALSO
CC EXPRESSED IN FAT BODIES. ON ENTRY INTO METAMORPHOSIS LEVELS OF
CC FULL LENGTH PROTEIN INCREASE IN TESTIS AND IMAGINAL DISKS.
CC -1- DEVELOPMENTAL STAGE: LARVAE, PUPAE AND ADULTS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC -1- CAUTION: REF.3 BELIEVES RESIDUE 690 IS A SELENOCYSTEINE.
CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous
CC gene model prediction.
CC -----
CC THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -
CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON
CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC ENTITIES REQUIRES A LICENSE AGREEMENT (See <http://www.1sb-sib.ch/announcement/>).
CC or send an email to license@1sb-sib.ch).
DR EMBL; L08483; AAA53471.1; -;
DR EMBL; L08483; AAA53472.2; -;
DR EMBL; AEO03657; AAF53651.1; ALT SEQ.

DR HSSP; Q05516; 1CS3.
DR FLYbase; Fggn0001301; kel.
DR InterPro; IPR000210; BTB POZ.
DR InterPro; IPR001798; Kelch.
DR Pfam; PR00651; BTB; 1.
DR Pfam; PR01344; Kelch; 6.
DR PRINTS; PR00501; KELCHREPEAT.
DR SMART; SM00225; BTB; 1.
DR PROSITE; PSS0097; BTB; 1.
DR Cytokeleton; Actin-binding; Selenium; Selenocysteine; Repeat.
KW Cytokeleton; Actin-binding; Selenium; Selenocysteine; Repeat.
FT CHAIN 1 1477
FT DOMAIN 1 689
FT REPEAT 157 223
FT REPEAT 404 449
FT REPEAT 450 496
FT REPEAT 498 543
FT REPEAT 545 592
FT REPEAT 594 639
FT REPEAT 641 687
FT DOMAIN 18 28
FT DOMAIN 29 36
FT DOMAIN 78 83
FT SE_CYS 690 690
FT SE_CYS 493 493
FT CONFLICT 596 596
FT CONFLICT 824 824
FT CONFLICT 858 858
FT CONFLICT 1083 1083
FT CONFLICT 1086 1086
SQ SEQUENCE 1477 AA; 160086 MW; 4851EAE9D9DBA47 CRC64;

Query Match 17.3%; Score 569; DB 1; Length 1477;

Best Local Similarity 27.3%; Pred. No. 1, 2e-13;

Matches 189; Conservative 104; Mismatches 249; Indels 98; Gaps 22;

DR 10 NTEYAVSLLEQLKFEYQQLFTDIVLIVEGTEPCHKMYATCGSYFRAMFSGLSSEKQ 69
DB 136 NQOHTARSPDANEMKOKOLCVILVADVEIHAHRMVASCSFYFAMFSGLSSEKQ 194
QY 70 THVHLRVNDAATLQIIITAYATGNLAMDSTVEQLYETACFLOVEVQLQRCREYLKIKN 129
DB 195 ARITLQSVARALELLIDVYATVAVNEDNVQLLTANLQLTVDRACCDFLTQTD 254
QY 130 AENCVALLSFADLFSCCEELKQSAKRWKFTAVYHQAEMOSSHLLIDISSDLNTE 189
DB 255 ASNCGLRIRFPADIHACVELLNAYETIEQHFENVIOPEDETLNLSHQVSLIGNDKISVP 314
QY 190 KEETVREAAMLLENTESRSQYLSVLSQIRIDALSE--VTQRAWFOGLPNDKSVVQ 247
DB 315 NEERVYECVIAWLRVDPWREQFTSLMEHVRLPFLSKVEYTORV-----DKELILB 366
QY 248 G-----LYSMKPKFK-----PR--LGMTKEEMFIEASSNPCLSYSSVCY 288
DB 367 GNIVCKNLITLALTYLLPLETSKARTVPKPKPMKPKPKPKPKPKPKPKPKPKPKPKPK 422
QY 289 SPOAEKYVLCSPPADLHKVGVTVTPDNDIYIAGGVQPLINTKNSKTSKIQATARPTN 348
DB 423 DLREKRYQAEMPNRRCRSGLSGLDCK-VYAVG-----FNSS-----LRVITVD 467
QY 349 CFYWPDAQONTWPKPTMLFVRIRKPSLVCEGYIYVAGISVSGELNRRTERVYDTEKBE 408
DB 468 V---YDPATDOWANGCNMEARSTLGVAVLNGCIVAGG--FDGTTGLSAAEMVDKTDI 522
QY 409 YTMVBLPCAMQMSAAVYVYHDCIYV-----TLNLYCYCFPPSDSVTEMAKQTSR 459
DB 523 MRPIASMSRRSSVGVVHGLIYVAGVDFTRQCLSSVERNPDTDMVNVVAEMSSRR 582
QY 460 SFASAAAFGDKIYIGGLHATNSGIRLPGCTVDGSSV--TVEIYVKNKEMQMANIYA 517
DB 583 SGAGVGLNIIYAVGG-H-----DGEVNRISVAICETNSMRSVADMSY 627
QY 518 KRYSDCVRVAVVISNS--LCVPMRETHLNERAKYVTVYQYDLRLDRWSLRHISERVLMWL 575

DB 628 CRRN-----AGVAVADGILYVVGDDGTNSNLASVEVYCPD--SDSMRLP-----ALMTI 675
QY 576 GRDFR--CTVGLKPSCLSE 593
DB 676 GRSTYAGVCKWIDK--DWCMEE 693

RESULT 3

KHLX HUMAN
ID Q9Y2M5; Q9H457; STANDARD; PRT; 609 AA.

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kelch-like protein X.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Yoshida K.; Sugano S.;
RT "Kelch motif containing protein."
RL Submitted (Apr-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE OF 9-609 FROM N.A.
RA Frankland J.;
RL Submitted (Sep-2000) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----

DR EMBL; AB026190; BA77027.1; -
DR EMBL; AL109921; CAC10469.1; -
DR InterPro; IPR000210; BTB POZ.
DR InterPro; IPR001798; Kelch.
DR Pfam; PR00651; BTB; 1.
DR Pfam; PR01344; Kelch; 6.
DR SMART; SM00501; KELCHREPEAT.
DR PRINTS; PR00501; BTB; 1.
DR PROSITE; PSS0097; BTB; 1.
KW Repeat.
FT DOMAIN 68 135
FT REPEAT 319 365
FT REPEAT 367 413
FT REPEAT 414 460
FT REPEAT 462 507
FT REPEAT 509 554
FT REPEAT 556 601
FT CONFLICT 593 593
SQ SEQUENCE 609 AA; 68083 MW; 67B2E3FE8341B422 CRC64;

Query Match 17.2%; Score 565.5; DB 1; Length 609;

Best Local Similarity 26.3%; Pred. No. 6, 6e-34;

Matches 151; Conservative 109; Mismatches 222; Indels 93; Gaps 14;

QY 9 INTEYAVSLLEQLKFEYQQLFTDIVLIVEGTEPCHKMYATCGSYFRAMFSGLSSEK 68
DB 46 ISDKPRLQTLVYINILRKRELCDVVLVGAKKIYAHRIYLSACSPYFRAMFTGELASR 105
QY 69 QTHVHLRVNDAATLQIIITAYATGNLAMDSTVEQLYETACFLOVEVQLQRCREYLKIKI 128
DB 106 QTEVIRIDIDERAMELLIDFATYSQITVEEGNVQTLPAACILQLAIEOACCEFLKQQL 165

QY 129 NAENCVRLSPADLFSCCELKQSAKRWVHEKFTAVYHQAQFMQLSHDLIDILSDNLVY 188
 Db 166 DPNCCGIRAFADTHSRRELLRIADKFTQNPFOEWESEFLLPANQLIDIISSDELVY 225
 QY 189 EKEETVREAMLMLEVTESRSQYLSVLSQIRIDALS-----E 227
 Db 226 RSEBOVFNAWAVAKVSIQERRPOLPQVLOHVRPLPLSPFLVGVGSDPLIKSDECRD 285
 QY 228 VTQRAMQGLPNDKSVVVGGLVYKSNPKFPRKPMGKEMMIPIEASSENPCS---LYS 284
 Db 286 LVEBAKRYLLLPQERPLMOG-----PR-TREPKPIGCVLAVAGW---CSGDAISS 334
 QY 285 SVCYSPQAEKRYKLCSPPADLHKVGTVPDNDIYAGS-----QVLEKNTKTNH-- 334
 Db 335 VERRYDPTQNMWVAVASKRCGVGVVL-DLLLYVAGHGDSSYLSNVRYDPTKNQMS 393
 QY 335 ----SKTSKLTQ-----FRTVNCYMDAQNTWFPPTPLPVRIKP 373
 Db 394 SDVAPSTCTRTSYGVAVLGFLYAVGQDVSCLNTVERYPKRNKTRVASMSTRRLGV 453
 QY 374 SLVCEGYIYAGDSVSGELNRTVERYPTEKDEMTWSPPLPCAMQMSAAVVVHDCIYV 433
 Db 454 AVAVLGGLYAVGSDSTSPLN--TVERYPQENRHTTIPMGTGRKHLGCAYQDVITA 511
 QY 434 M-----TLNLMYCYFPRSDSWEMAMRQTSRSPASAAFGDKIFYIGLHIAITNSGAR 486
 Db 512 VGGRDPTTELSAERYVPRTNQMSPVYAMTSRSRGVGLAVVNGQLMAVG-----562
 QY 487 LPSGYDGSV--TVEIYDVKNEKMAANIIPAKR 519
 Db 563 -----DGTLYLKTIEVDPDANTWRLYGGMYNR 591

RESULT 4

KRPL_HUMAN STANDARD; PRT; 606 AA.
 AC 060662;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Kelch-related protein 1 (Kel-like protein 23) (Sarcosin).
 GN KRPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (SHORT ISOFORM).
 RX MEDLINE=98317890; PubMed=9655184;
 RA Taylor A., Obholz K., Linden G., Sadlev S., Klaus S., Carlson K.D.;
 RT "DNA sequence and muscle-specific expression of human sarcosin
 transcript";
 RT Mol. Cell. Biochem. 183:103-110(1998).
 RN [2]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RX TISSUE=Skeletal muscle;
 RX MEDLINE=20180502; PubMed=10713668;
 RA Spence H.J., Johnston I.P., Ewart K., Buchanan S.J., Fitzgerald U.,
 Ozanne B.W.;
 RT "Krlp, a novel kelch related protein that is involved in pseudopod
 elongation in transformed cells";
 RT Oncogene 19:1266-1276(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (LONG ISOFORM).
 RA Zhuang D., Gunnarsson D., Toffia O., Lind M., Lundgren P.,
 Selstam G.;
 RT "Mammalian Kel or Kel-like proteins related to ovarian development and
 differentiation";
 RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR PSEUDOPOD ELONGATION IN TRANSFORMED CELLS.
 CC -1- SUBCELLULAR LOCATION: PREDOMINANTLY CYTOSOLIC BUT CAN CO-
 LOCALIZE WITH F-ACTIN AT THE MEMBRANE RUFFLE-LIKE STRUCTURES AT
 THE TIPS OF TRANSFORMATION-SPECIFIC PSEUDOPODIA.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: SARCOMERIC MUSCLE.
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 KELCH REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL: AF056929; AAC13686.1; -
 CC EMBL: AF333387; AAC52886.1; -
 CC InterPro: IPR000210; BTB_POZ.
 CC Pfam: PF00651; BTB; 1.
 CC Pfam: PF01344; Kelch; 5.
 CC SMART: SM00225; BTB; 1.
 CC PROSITE: PS50097; BTB; 1.
 CC CyscoSkelton; Repeat; Alternative splicing.
 KM DOMAIN 33 100
 FT REPEAT 346 398
 FT REPEAT 399 447
 FT REPEAT 448 495
 FT REPEAT 497 542
 FT REPEAT 544 599
 FT VARSPLIC 1 11
 FT CONFLICT 304 304
 FT CONFLICT 358 358
 FT CONFLICT 361 361
 SO SEQUENCE 606 AA; 68037 MW; 8C7BC13EBE601034 CRC64;
 Query Match 16.8%; Score 554; DB 1; Length 606;
 Best Local Similarity 26.6%; Pred. No. 4,6e-33;
 Matches 155; Conservative 93; Mismatches 240; Indels 94; Gaps 13;

QY 6 EQQINTE---YAVSLLEQ-LKIFYEQQLFTDVLVIVEGTFEPCHEMVAATSSSYPRAMW 61
 Db 4 QRELAEELRYOSTLLQDGLKDLDBKEFTDCTLAGDKSLCHRLITISACSPYREFEYL 63
 QY 62 SLSESQTHVLRNVDAATLQIIITYATGNAMNDSTVEOLYETACFLYQVEDYLQRCR 121
 Db 64 SEIDAKKEVLDNDVPAIDLLIILIKYLSASIDNDGVQDIFALASRQIPSVFTVCV 123
 QY 122 EYLKIKINAEVCVRLSPADLFSCCELKQSAKRWVHEKFTAVYHQAQFMQLSHDLIDIL 181
 Db 124 SYLQKRLAPGNCIALIRGLILDCPLAISAREFVSDRFQICKEDFMQLSPQELISVI 183
 QY 182 SSDNLNVEKEETVREAMLMLEVTESRSQYLSVLSQIRIDALS-----227
 Db 184 SNDLSNVEKEEAVFERVAVKVRITDKENRKNLSEVFDCLRFLMTKEKFKDHEKDDIIL 243
 QY 228 -----VTQRAMQGLPNDKSVVVGGLYK-----SNPKFKP--PRLGMTKE 266
 Db 244 SNPDQKKIKVLDKAPAGLTPSPSKNAKKTAGEVNGVDGDDLLPGYINDIPRGMFVK 303
 QY 267 EMMIFIEASSENPCSLYSSVCSYPOAEKRYKLCSPPADLHKVGTVPDNDIYIAGQVP 326
 Db 304 DILLVNDT-----AAVAYDPTENECYLTALAEQIPRNHSSIVTQONQIYVVGGLY 354
 QY 327 LKNTKTNHKSQKLTQAPRTVACFWPDAQQNTWFPKTPMLFVRIRKPSLVCEGY-----381
 Db 355 -----VDEENKQDPLD-----SYFQDLSIASEWGLPEPL-----PSARCLFGGEVVD 398
 QY 382 -IYAGDSVSGELNRTVERYPTEKDEMTWSPPLPCAMQMSAAVVVHDCIYV-----434
 Db 399 KIYVAGDQLQELASDLSVLTCDPVAAKNEVKGLPIKYGINNVSHKMITCLGKTD 458
 QY 435 --TLNLMYCYFPRSDSWEMAMRQTSRSPASAAFGDKIFYIGLHIAITNSGIRLPSGTV 492

OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99326431; PubMed=10397770;
 RA Soltyseik-Espanola M.B., Rogers R.A., Jiang S., Kim T.A., Gaedigk R.,
 RT "Characterization of Mayven, a novel actin-binding protein
 "predominantly expressed in brain."
 RL Mol. Biol. Cell 10:2361-2375(1999).
 CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF
 THE BRAIN CELLS.
 CC -1- SUBUNIT: BINDS ACTIN.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN.
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 6 KETCH REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF059569; AAC67502.1; -.
 CC HSSP; O05516; 1CS3.
 CC Genew; HGNC:6353; KLMH2.
 CC MIM; 605774; -.
 CC InterPro; IPR000210; BTB_POZ.
 CC InterPro; IPR001798; Ketch.
 CC Pfam; PF00651; BTB; 1.
 CC Pfam; PF01344; Ketch; 6.
 CC PRINTS; PRO0501; KETCHREPEAT.
 CC SMART; SM00225; BTB; 1.
 CC PROSITE; PSS0097; BTB; 1.
 CC Cytokeleton; Actin-binding; Repeat.
 CC FM DOMAIN 56 123 BTB.
 CC REPEAT 308 353 KETCH 1.
 CC REPEAT 354 400 KETCH 2.
 CC REPEAT 402 447 KETCH 3.
 CC REPEAT 449 496 KETCH 4.
 CC REPEAT 497 543 KETCH 5.
 CC REPEAT 545 591 KETCH 6.
 CC SEQUENCE 593 AA; 66033 MW; 50479F44F359A81E CRC64;
 SQ
 Query Match 16.2%; Score 534; DB 1; Length 593;
 Best Local Similarity 25.3%; Pred. No. 1.3e-11;
 Matches 138; Conservative 100; Mismatches 199; Indels 108; Gaps 14;
 OY 5 DEROINTE-----YAVSLLEQLKLFYEOQLFTDIYLYEGTEFFPCHKVLA 51
 DB 20 DSKDNDTEKHCPTVTPVPMHMKAFKVMNELR---SQNLCDVTIVADEMISHRVYLA 76
 OY 52 CSSYFPAMFGSGSEKQTHVLRNDATLQIITYATGNLANDSTYEQLYTACPL 111
 DB 77 CSFYFHAMFGEMSESRKVRKEVDGWTIRMLIDYVYVAEIQVEENQVLLPAGLL 136
 OY 112 QVDDVLCREYLIKINMNCRLLSPADLFSCEELKQSAKRMVHKFAVYVHQAQFMQ 171
 DB 137 QLODVAKTCEEFESQHPVNCIGIRFADMAHACTDLNKANTYAOHFADVLLSEEFN 196
 OY 172 LSHDLIDLSNINLWKEKETREANMLLEVNTESRSQYSSVLSQIRIDALSE--VT 229
 DB 197 LGIEQYCSLISSKLTISSEKVFNAVIMVNDKVRQFMARLMEHRYLLPLREYLV 256
 OY 230 QRAMFGGLPND--KSVVQGL-YSMKPEFFRLGKMTKEEMIFTEASSENPCSLYS 285
 DB 257 QRVEEBALVNSACKRYLEIAKKYLLP-----TEORILM----- 292
 OY 286 VCVSPQAEKYYKLCSPADLHKGTVTTPNDIYIAGQVPLKNTNNSKSKQTARF 345
 DB 293 -----KSVRTLRTPMNLKLMVVV-----GGQAP-----KAIR 321

OY 346 TVNCFYFDAQNTWPEKTPMLFVRIKPSLVCCGEYIYAIGDSDVGGELNRRTERYTE 405
 DB 322 SVCE---YDFKEBRHWQVAELPSRRRCRAGMYMAGLVFVAVG--FNGSLRVRTSDSYDV 376
 OY 406 KPEWTVSPPLPCAMQMSAAVYVHDCIYV-----TLNLMCYFPRRSWYEMMRQTS 458
 DB 377 KOQWTSVANMRDRSTLGAIVANGLLYAVGFGDSTGLSSVAYNINKSEWPHVAMNTR 436
 OY 459 RSPASNAFQDKIFYIGLHINTSGIRLPSGTVDGS---VVEIYDVNKNWMAAN 514
 DB 437 RSVGVGVGGLLVAVG-----YQASQYVSTJECVATINNEWYIAE 481
 OY 515 IPAKR 519
 DB 482 MSTRR 486
 RESULT 7
 ID KHL1 HUMAN STANDARD; PRT; 748 AA.
 AC Q9NR64; Q9NR65; Q9P238; Q9H4X4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Ketch-like protein 1.
 GN KHL1 OR KIAA1490.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20347694; PubMed=1088605;
 RA Koob M.D., Nemes J.P., Benzow K.A.;
 RT "The SCA8 transcript is an antisense RNA to a brain-specific
 RT transcript encoding a novel actin-binding protein (KHL1).";
 RL Hum. Mol. Genet. 9:1543-1551(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20277482; PubMed=10819331;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirotsawa M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 7:143-150(2000).
 RN [3]
 RP SEQUENCE OF 179-409 FROM N.A.
 RA Kay M.;
 RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF
 THE BRAIN CELLS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN.
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 6 KETCH REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF252283; AAF81719.1; -.
 CC EMBL; AF252279; AAF81716.1; -.
 CC EMBL; AB040923; BAA96014.1; ALT_INIT.
 CC EMBL; AL353738; CAC16128.1; -.
 CC Genew; HGNC:6352; KHL1.
 CC MIM; 605332; -.
 CC InterPro; IPR000210; BTB_POZ.
 CC InterPro; IPR001798; Ketch.

```

DR pEam; PF00651; BTB; 1.
DR pEam; PF01344; Kelch; 6.
DR PRINTS; PR00501; KELCHREPEAT.
DR SMART; SMO0225; BTB; 1.
DR PROSITE; PSS0097; BTB; 1.
KM Cytoskeleton; Actin-binding; Repeat.
PT DOMAIN 43 88
PT REPEAT 212 279 BTB.
PT REPEAT 460 506 KELCH 1.
PT REPEAT 507 553 KELCH 2.
PT REPEAT 555 600 KELCH 3.
PT REPEAT 601 647 KELCH 4.
PT REPEAT 649 700 KELCH 5.
PT REPEAT 701 747 KELCH 6.
SQ SEQUENCE 748 AA; 82680 MW; C11C43D8282F9F9 CRC64;

Query Match 15.7%; Score 523; DB 1; Length 748;
Best Local Similarity 25.0%; Pred. No. 1,1e-30;
Matches 151; Conservative 101; Mismatches 237; Indels 114; Gaps 15;

2 STODERQITVEYAVSLLEQLKLFYEQQLFTDIYLYEGTEPFCMKVLTCSYFPAFM 61
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 SSSSEYQAVHNAHQFPRKKESTLKQOQLCDVLIYGNKRIHKLVSVDYFPAFM 242
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 SGLSEKQTHVHLRNVDATLQIITAYATGNLAMDSTVEQLYETACFLQVEDVLRGR 121
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
243 SDVCEAKQEKIKMEGIDPNALMDLVQFAATGCLELKEDFTENLAAACLLQLPQVEVCC 302
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
122 EYLKIKINENCYRRLSPADLFSCBELKQSAKMWENKFAVYVHQAQFQSLDILIDL 181
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
303 HFLMKLHPNSNCIGIAPADQCCILMKVAHSYTMENIWEIVIRNOEFLILPKEELHKL 362
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
182 SSDNLVKEETVREAAMLLEXYNTESRQYSSVLSQIRI-----DALSEVTRAMFOG 236
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
363 ASDDVAVPBEETIFHLMMWVKYDKMSRCNDLSMLAFIRLLPQLADLENHALLFKK 422
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
237 LEPNDKSVVVOGLYKSMK-----FKKPR----- 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
423 DLEQQLILEAMKYLHLLPERRILMQSPRTKRSVTGTLVAVGMDNNKGATTIEKYDR 482
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
261 -----LGMTKEEMTIFIESSNPCLSVSYCSQAKEYVGLCSPPAD 304
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
483 TNLMTQAGMNGRRLOFGVAVIDDKLFVIGRRGLKTLNIVCEYNPKTKTWVL--PPMS 540
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
305 LHKVGTVTTP-DNDIYIAGQVPLKTKTNHSTKSLQTFRTVNGCFYVMDAQNTWPK 363
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
541 TRRHGIGVTVLESPITAVGG-----HDSWSTLNTVER-----WDPOSQMTYV 583
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
364 TPLFVRIKPSLVSCGCIYIAGDSVGGELNRTVERVYTEKDEWTNVSPLPCAMQWSA 423
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
584 ASMSIARSTVGVAAALNGKLYSGVRD--GSSCLSSMEYVPHNNKNNMCAPM-CKRRGV 640
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
424 AVYVND-CIYVM-----TLNLMYCYFPSPDSVVENAKQTSRSPASAAAFGD 469
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
641 GVAATCGFLYAVAGHDAPASNHCRLDYVERDPRTDTVTWVAFLSPMPDAVGVCLGD 700
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
470 KIFYGGLHATNSGIRLPSGTWDGSSV--TVEIVDNKMEGMKANIPAKRSDPCVRA 527
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
701 RLTVAGG-----YDQOTYANTMESYDPQNTMEYQMASLNGR-AGACV-- 742
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 528 VVI 530
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 743 VVI 745
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
ENCL MOUSE STANDARD; PRT; 589 AA.
AC 035709;
DT 16-OCT-2001 (Rel. 40; Created)
DT 16-OCT-2001 (Rel. 40; Last sequence update)
DT 15-UTN-2002 (Rel. 41; Last annotation update)
DR Ectoderm-neural cortex-1 protein (ENC-1).

```

```

GN ENCL OR ENC-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Swiss albino; TISSUE=Brain;
RX MEDLINE=97252647; PubMed=9096139;
RA Hernandez M.-C., Andres-Barquin P.J., Martinez S., Bulfone A.,
RA Rubenstein J.L.R., Israel M.A.
RT "ENC-1: a novel mammalian kelch-related gene specifically expressed in
RT the nervous system encodes an actin-binding protein."
RL J. Neurosci. 17:3038-3051(1997).
CC -1- FUNCTION: ACTIN-BINDING PROTEIN INVOLVED IN THE REGULATION OF
CC NEURONAL PROCESS FORMATION AND IN DIFFERENTIATION OF NEURAL CREST
CC CELLS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. INTERACTS WITH THE ACTIN
CC CYTOSKELETON.
CC -1- TISSUE SPECIFICITY: PRIMARILY EXPRESSED IN THE NERVOUS SYSTEM.
CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS HIGHLY DYNAMIC BUT MOSTLY
CC RESTRICTED TO THE NS. OUTSIDE THE NS, EXPRESSION IS DETECTED IN
CC THE ROSTRAL-MOST SOMITOMERE OF THE PRESOMITIC MESODERM, AT THE
CC TIMES CORRESPONDING TO THE EPITHELIALIZATION THAT PRECEDES SOMITE
CC FORMATION. FIRST DETECTED IN THE BRAIN AND SPINAL CHORD OF 12 PC
CC EMBRYOS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; U65079; AAB64206.1; -.
CC MGD; MG1:109610; Enc1.
CC DR InterPro; IPR000210; BTB_POZ.
CC DR InterPro; IPR001798; Kelch.
CC DR Pfam; PF00651; BTB; 1.
CC DR Pfam; PF01344; Kelch; 5.
CC DR SMART; SMO0225; BTB; 1.
CC DR PROSITE; PSS0097; BTB; 1.
CC KM Actin-binding: Developmental protein; Cytoskeleton; Repeat.
CC PT DOMAIN 46 114
CC PT REPEAT 296 340
CC FT REPEAT 341 388
CC FT REPEAT 389 444
CC FT REPEAT 446 492
CC FT REPEAT 494 538
CC FT REPEAT 539 585
CC FT REPEAT 589 642
CC SQ SEQUENCE 589 AA; 66085 MW; 12B62354D508B6A2 CRC64;

Query Match 15.7%; Score 516.5; DB 1; Length 589;
Best Local Similarity 25.5%; Pred. No. 2,5e-30;
Matches 155; Conservative 94; Mismatches 237; Indels 122; Gaps 17;

11 TEYAVSLLEQLKLFYEQQLFTDIYLYEGTEPFCMKVLTCSYFPAFMGSLSSKOT 70
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
26 SSVADSVLTHNLNLKQORLFTDVLHAGNRTFPCRAVLAACSRFEAMFSGLKESQDS 85
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
71 HYLNR-VDAATLQIITAYATGNLAMDSTVEQLYETACFLQVEDVLRGREYLYIKIN 129
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
86 EVNFDNSHPEVLELLDVAYSRVYINENASLLEAGDMLEFQDIRDACAPELEKNIH 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
130 AENCYRLSPADLFSCBELKQSAKMWENKFAVYVHQAQFQSLDILIDLISDNLNAYE 189
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
146 PNLCLGMLLSDAHQCTKLYELSWMCNFPQTIKSEDFLOLPQDMVVQGLSSBELFTE 205
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
190 KEETVREAAMLLEXYNTESRQYSSVLSQIRIDALSEVTRAMFOGLPNDKSVVVOGL 249
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 206 DERLYVESAMNWSYDLKKRYCYLPELLQTVRLALLPAIYIM-----ENVAMEEL 255
 QY 250 YKSMPEKFKPRLGWTKE--EMMTFIEASSENPCSLYSVCYSPQ-----A 292
 Db 256 ITQOKR-----SKEIYEAIRCKKILIONGVSTSLCARPKTGHALFLAGGQTMG 307
 QY 293 EKVY-----KLGSPPADL--HKVGVTVTPDDIYIAGOVPLKNTKTHSKTSKLQTA 343
 Db 308 DKLYVDQKAKEIIPKADIDSPRKEFSAICAGKVVYITGG--RSENGVSKD----- 357
 QY 344 FRTVNCYCW-FDAQONTWPFKTPMLFRTFRLPSLVCCGGIYIYAGDS-----VGGELN 395
 Db 358 -----VWVYDTLHEEWSKAAPMLVARFGHGAELKGLYVVGHTAATGCLPASPSS 410
 QY 396 RRTVERYDEKDEMTWSPPLPCAMQMSAAV-----VHDCIYVMTLIMLYCY 442
 Db 411 LKQVEQVDPTTKNTKTVAPILREGVSNAAYSAKLKLPAGGTSVSHD-----KLPKYQCY 465
 QY 443 FPRSDSVWEMAMQTSRSPASAAAFGDKIFYIGLHIAATNSGIRLPSGTVDGSSVTEIY 502
 Db 466 DGCENRWSVPATCPQPPRYTAAAVLGNQIFIMG-----GDTEFSAGSAYKF 511
 QY 503 DVNKNEMKMAANIIPAKRYSPCRAVAVTNSLCYPRKETHLNRKATVYQ-----YD 555
 Db 512 NSSETYQTKVGDVYAKRMS--CHAAVSGNKLYV-----GGYFGIQRCKTIDCYD 559
 QY 556 LEIDRWS 562
 Db 560 PTLIDVWN 566

RESULT 9
 KHL1 MOUSE STANDARD; PRT; 751 AA.

AC 09J174;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Kelch-like protein 1.
 GN KHLHL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;
 NC NCB1 TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20347694; PubMed=10888605;
 RA Koob M.D., Nemes J.P., Benzon K.A.;
 RT "The SCa8 transcript is an antisense RNA to a brain-specific
 transcript encoding a novel actin-binding protein (KHLHL1).";
 RL Hum. Mol. Genet. 9:1543-1551(2000).
 CC -1- FUNCTION: MAY PLAY A ROLE IN ORGANIZING THE ACTIN CYTOSKELETON OF
 THE BRAIN CELLS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN BRAIN.
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation-
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AF252281; AAF81717.1;
 DR MGD; MGJ:2136335; K1h11.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF00651; BTB_1.
 DR Pfam; PF01344; Kelch; 6.
 DR PRINTS; PR00501; KELCHREPEAT.

DR SMART; SM00225; BTB; 1.
 DR PROSITE; PS50097; BTB; 1.
 KW Cytoskeleton; Actin-binding; Repeat.
 FT DOMAIN 45 90
 FT BTB.
 FT 215 282
 FT 463 509
 FT REPEAT 510 556
 FT REPEAT 558 603
 FT REPEAT 604 650
 FT REPEAT 652 703
 FT REPEAT 704 750
 FT REPEAT 751 AA; 82932 MW; 6E846190CAC6B4C3 CRC64;
 SQ SEQUENCE

Query Match 15.7%; Score 516; DB 1; Length 751;
 Best local similarity 24.9%; Pred. No. 3,7e-30;
 Matches 150; Conservative 101; Mismatches 238; Indels 114; Gaps 15;

QY 2 STQDERQINTEXAVSLLEQLKLPYEQQLFTDYLIVLEGTEFPCHKVTLATCSSYFRAMFM 61
 Db 186 SSSEEFYQAVRHAEQGFRRKENVYLKQOOLCDVILLIGNKKIPAHRLVLSVSDYFAAMFT 245
 QY 62 SGLSESKQTHVHLRNVDATLQIITAYATGULAMNDSTVEQLYETACLOVEDVLCQR 121
 Db 246 SDVCEAKQKEITKMEGIDPNALMDLVQFATGCLCELEKEDTLENLAACLDLPQVEVCC 305
 QY 122 EYLKIKINENCVRLLSFADLFCSEELKQSAKRWENKHTAAVYHODAFMQLSHDLIDL 181
 Db 306 HFLMKLHPSNCLGIRAFADAGCIELMKVASYMENMEVIRNGEFLFLPAEBELKHL 365
 QY 182 SSDNLNVEKEFTVREAAIMLEYNTEBSRQYLSVLSQIRI-----DALEVTQRAVFG 236
 Db 366 ASDDVAVPDEETIFHALMMWVYKDMQRRCSDSLMLAFYRLDLPQIILADLENHALFYK 425
 QY 237 LEPNDKSVVVGGLYSMPK-----FFKPR----- 260
 Db 426 DLBCQKLLEAMKTHLLPERRTLMOSPTKPKSVTGILYAVGMDNKGATTEIKYDRL 485
 QY 261 -----LGWTKEMMIFIEASSENPCSLYSVCYSPQAEKVKYLCSPAD 304
 Db 486 TNLMTIQAGMNGRRRLQFGAVVIDDKLFLVIGRSGDKLTALTVEGCYNPKTKTWTVL--PMS 543
 QY 305 LHKVGTVTP-DNDIYIAGQVPLKNTKTHSKTSKLOTAFTVNCVCFDQAQONTWPK 363
 Db 544 TRHIGLGVTVLGGPIYAVG-----HDGWSYLNATVER-----WDPOSQQWTYV 586
 QY 364 TMLLFRIRPSLVCCGGIYIATGDSVGGELNRRTYRYDEKDEMTWSPPLPCAMQMSA 423
 Db 587 ASMSIARSTVGVAALNGKLYSVGRD--GSSCLSSMEYTDPTTKMSKCPM--CKKRGV 643
 QY 424 AVVVDH-CIYV-----TLNLMCYFPRSDSVWEMAMQTSRSPASAAAFGD 469
 Db 644 GVAICDGFPLVAVGHDAPASNNCSRLDVEVEYEPRTDVTWVAPLSPMRDAVGVCLLD 703
 QY 470 KIFYIGLHIAATNSGIRLPSGTVDGSSV--TYEIVDNKEMKMAANIIPAKRYSPCRA 527
 Db 704 RLYAVG-----YDQTYLNTWESYDPTQNTWOMASINIGR-AGACV-- 745
 QY 528 VVI 530
 Db 746 VVI 748

RESULT 10
 ENCL1 HUMAN STANDARD; PRT; 589 AA.

AC O14682; Q9UPG9; O75464;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Ectoderm-neural cortex-1 protein (ENC-1) (p53-induced protein 10)
 DE (Nuclear matrix protein NRP/B).
 GN ENCL1 OR P1010 OR NRPB.
 OS Homo sapiens (Human).

OS	KLHL4 OR KIAA1687.	
OC	Homo sapiens (Human).	
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;	
CC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
OX	NCBI_TaxId:9606;	
RN	[1]	
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).	
RX	MEDLINE=21295031; PubMed=11401425;	
RA	Brybrook C., Warty G., Howell G., Anasson A., Bjornsson A.,	
RA	Moore G.E., Ross M.T., Standler P.;	
RT	"Identification and characterization of KLHL4, a novel human homologue	
RT	of the Drosophila Kelch gene that maps within the X-linked cleft	
RT	palate and ankyloglossia (CPX) critical region.";	
RL	Genomics 72:128-136(2001).	
RN	[2]	
RP	SEQUENCE FROM N.A. (ISOFORM 1).	
RC	TISSUE=Brain;	
RX	MEDLINE=21082932; PubMed=11214970;	
RA	Nagase T., Kikuno R., Hattori A., Kondo Y., Okumura K., Ohara O.;	
RT	"Prediction of the coding sequences of unidentified human genes. XIX.	
RT	The complete sequences of 100 new cDNA clones from brain which code	
RT	for large proteins in vitro."	
RL	DNA Res. 7:347-355(2000).	
RN	[3]	
RP	SEQUENCE OF 133-718 FROM N.A. (ISOFORM 1).	
RA	Itoigai T., Oca T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,	
RA	Nishikawa T., Nagai K., Sugano S., Shiyama T., Sudo H.,	
RA	Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo Y., Sugawara M.,	
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,	
RA	Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,	
RA	Yanamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,	
RA	Ninomiya K., Iwayanagi T.;	
RT	"NDO human cDNA sequencing project."	
RT	Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.	
RL	[4]	
RP	SEQUENCE OF 150-718 FROM N.A. (ISOFORM 1).	
RA	Bird C.;	
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.	
CC	-1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2/KLHL4c; are	
CC	produced by alternative splicing.	
CC	-1- TISSUE SPECIFICITY: Expressed in adult fibroblasts and in a range	
CC	of fetal tissues including tongue, palate, and mandible.	
CC	-1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.	
CC	-1- SIMILARITY: CONTAINS 6 KELCH REPEATS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AF284765; AAK9441.1; -	
DR	EMBL; AF284766; AAK9442.1; -	
DR	EMBL; AB051474; BAB21778.1; ALT_INIT.	
DR	EMBL; AK022715; BAB14199.1; ALT_INIT.	
DR	EMBL; AL035424; CAB39994.1; -	
DR	Genew; HGNC:6355; KLHL4.	
DR	MIM; 300348; -	
DR	InterPro; IPR000210; BTB_POZ.	
DR	InterPro; IPR001798; Kelch.	
DR	Pfam; PF00651; BTB; 1.	
DR	Pfam; PF01344; Kelch; 6.	
DR	PRINTS; PR00501; KELCHREPEAT.	
DR	SMART; SM00225; BTB; 1.	
DR	PROSITE; PS50097; BTB; 1.	
KW	Cyoskeleton; Actin-binding; Repeat; Alternative splicing.	
FT	DOMAIN 182 249	
FT	REPEAT 430 476	
FT	REPEAT 477 523	
FT	REPEAT 525 570	
FT	REPEAT 571 617	

FT	REPEAT	619	670	KEICH 5.
FT	REPEAT	671	717	KEICH 6.
FT	REPEAT	671	717	KEICH 6.
FT	VARSPLIC	700	718	EVFNPIGKAGACVYVVKLP -> SMOELLQNFYTTQKLET
FT	SEQUENCE	718 AA;	80245 MM;	LGH (IN ISOFORM 2).
FT	SEQUENCE	718 AA;	80245 MM;	D7385BAAD63B150A CRC64;
Query Match		15.4%;	Score 505.5;	DB 1; Length 718;
Best Local Similarity		24.6%;	Pred. No. 2e-29;	
Matches 148;	Conservative 103;	Mismatches 233;	Indels 111;	Gaps 17
QY	2	STODERQINT-EYAVSLLEQLKLEYEQOLFPTDIVLIVBETFPCHKVLATCSSYFRAMF	60	
DB	152	ATREBEQPHVINHAEQTLRKMEVNLKEKQCLDVLILAGHLNIPARHLVLSAVSDYFAAMF	211	
QY	61	MSGISESKQTHVHLRNDAATLQIIITTYATYGNLAMDSTYEQLYETKCFIQVEDYLORC	120	
DB	212	TNDVLEAKQOEVEVRMEGVDPNALNSLVQYAAVYGVQLKEDITESLAAACLLDQLTQVIDVC	271	
QY	121	REYLIIKKINANCRLISFADLPFCSEELKQSAKAMVHEKFPNAVHODAFMQLSHLDLDI	180	
DB	272	SNFLIKQOLHPNSCLGINSFGDAQCTELNVANHXYMEHFLVINKQEFFLLPAEISKL	331	
QY	181	LSNSLNVEKEETVREAAMLWLEVENTESRQYLSVLSQIRI-----DALSEVTORAMFQ	235	
DB	332	LCSDDINVPDEETTFHAIQVGVGHVDQVQROGELCMLSYIPLPLPQLADLETSSMFT	391	
QY	236	G-----LPPNDKSYV-----VQGLY-----KSNKPF--	256	
DB	392	GDLECOKLMEAMKYHLLPERSMMQSPRTKPRKSTVGALYAVAGMDAMKGTITTEKYDL	451	
QY	257	-----FKPRLGKTEKEMMFIEMASSENPGLSYSCVSPQAEKYVTKLCSRA	303	
DB	452	RTNSWMLHIGTNGRRLOPGAVINDNKLYVGGRGKLTNTVECPNP-VGKIWTY-MPPM	509	
QY	304	DLHKVGT-VTPDNDIYAGQVPLKNTKTNHSKTSKQTAFRYTNCFYEMDAQONTWFP	362	
DB	510	STHHGGLGVALTEBGMVAVG-----HDKMSYLTNYER-----WDPEGRQNY	552	
QY	363	KTPMLFVRIKPSLVCCBEGYIYAGDSVGBELNRYTERYDTEKEDWTVSPPLPCAWMS	422	
DB	553	VASNSTPRSTYGVVALNKKLYAIGRD--GSSCLKSMYFPDHTKWSLCLAPMSKRGGV	610	
QY	423	AAVVVHDCITYM-----TLNLWC--YFPRSSVWEMAMKQTSRSGFASAAAGD	469	
DB	611	GVATYNGFLYVVGHDAPASNHCSRSLSPCVERYPKQGSWSTVAPLSVPRPAVAVACPGLD	670	
QY	470	KIPIYIGGHLIATNSGIRLPSTGVGSSV--TVELYDVVKNMKMAANIPAKRYSDPCYRA	527	
DB	671	KLYVVG-----YDGHYTLNTVESYDAORNMKEEVPVNIGR-AGACVAV	714	
QY	528	V 528		
DB	715	V 715		
RESULT 12				
KHLS_HUMAN	STANDARD;	PRT;	755 AA.	
AC	Q96PQ7; OGNMUK3;			
DT	15-JUN-2002 (Rel. 41, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Keich-like protein 5.			
GN	KHLS.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Mao Y., Xie Y., Wang S., Zhou Z., Zhao W., Wang W., Zhao S.;			
RT	"Isolation and characterization of KHLS, a novel human gene encoding a keich-related protein with a B7B domain."			

```

RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 212-755 FROM N.A.
RC TISSUE=placenta;
RA Isogai T., Ota T., Hayaishi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagaitsuna M., Hoshita T., Kaku Y., Kodaera H., Kondo H., Sugawara M.,
RA Wakaishi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saio K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,
RA Niimiya K., Iwasanagi T.;
RT "NEO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
CC stop codon in position 735.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL, AF22976; AAL08584.1; ALT_SBO.
CC EMBL, AK002174; BAA92121.1; ALT_INTT.
CC Genew, HGNC:6356; KHLH5.
CC InterPro, IPR000210; BTB_POZ.
CC InterPro, IPR001798; Kelch.
CC Pfam, PF01344; Kelch; 5.
CC PROSITE, PSS0097; BTB; 1.
CC DR Pfam; P01344; Kelch; 5.
CC DR PROSITE; PSS0097; BTB; 1.
CC KW Cytoskeleton; Actin-binding; Repeat.
CC FT DOMAIN 220 287
CC FT REPEAT 468 514
CC FT REPEAT 515 561
CC FT REPEAT 563 608
CC FT REPEAT 609 655
CC FT REPEAT 657 708
CC FT REPEAT 709 754
CC FT REPEAT 754 835
CC FT REPEAT 835 895
CC FT REPEAT 895 955
CC FT REPEAT 955 1015
CC FT REPEAT 1015 1075
CC FT REPEAT 1075 1135
CC FT REPEAT 1135 1195
CC FT REPEAT 1195 1255
CC FT REPEAT 1255 1315
CC FT REPEAT 1315 1375
CC FT REPEAT 1375 1435
CC FT REPEAT 1435 1495
CC FT REPEAT 1495 1555
CC FT REPEAT 1555 1615
CC FT REPEAT 1615 1675
CC FT REPEAT 1675 1735
CC FT REPEAT 1735 1795
CC FT REPEAT 1795 1855
CC FT REPEAT 1855 1915
CC FT REPEAT 1915 1975
CC FT REPEAT 1975 2035
CC FT REPEAT 2035 2095
CC FT REPEAT 2095 2155
CC FT REPEAT 2155 2215
CC FT REPEAT 2215 2275
CC FT REPEAT 2275 2335
CC FT REPEAT 2335 2395
CC FT REPEAT 2395 2455
CC FT REPEAT 2455 2515
CC FT REPEAT 2515 2575
CC FT REPEAT 2575 2635
CC FT REPEAT 2635 2695
CC FT REPEAT 2695 2755
CC FT REPEAT 2755 2815
CC FT REPEAT 2815 2875
CC FT REPEAT 2875 2935
CC FT REPEAT 2935 2995
CC FT REPEAT 2995 3055
CC FT REPEAT 3055 3115
CC FT REPEAT 3115 3175
CC FT REPEAT 3175 3235
CC FT REPEAT 3235 3295
CC FT REPEAT 3295 3355
CC FT REPEAT 3355 3415
CC FT REPEAT 3415 3475
CC FT REPEAT 3475 3535
CC FT REPEAT 3535 3595
CC FT REPEAT 3595 3655
CC FT REPEAT 3655 3715
CC FT REPEAT 3715 3775
CC FT REPEAT 3775 3835
CC FT REPEAT 3835 3895
CC FT REPEAT 3895 3955
CC FT REPEAT 3955 4015
CC FT REPEAT 4015 4075
CC FT REPEAT 4075 4135
CC FT REPEAT 4135 4195
CC FT REPEAT 4195 4255
CC FT REPEAT 4255 4315
CC FT REPEAT 4315 4375
CC FT REPEAT 4375 4435
CC FT REPEAT 4435 4495
CC FT REPEAT 4495 4555
CC FT REPEAT 4555 4615
CC FT REPEAT 4615 4675
CC FT REPEAT 4675 4735
CC FT REPEAT 4735 4795
CC FT REPEAT 4795 4855
CC FT REPEAT 4855 4915
CC FT REPEAT 4915 4975
CC FT REPEAT 4975 5035
CC FT REPEAT 5035 5095
CC FT REPEAT 5095 5155
CC FT REPEAT 5155 5215
CC FT REPEAT 5215 5275
CC FT REPEAT 5275 5335
CC FT REPEAT 5335 5395
CC FT REPEAT 5395 5455
CC FT REPEAT 5455 5515
CC FT REPEAT 5515 5575
CC FT REPEAT 5575 5635
CC FT REPEAT 5635 5695
CC FT REPEAT 5695 5755
CC FT REPEAT 5755 5815
CC FT REPEAT 5815 5875
CC FT REPEAT 5875 5935
CC FT REPEAT 5935 5995
CC FT REPEAT 5995 6055
CC FT REPEAT 6055 6115
CC FT REPEAT 6115 6175
CC FT REPEAT 6175 6235
CC FT REPEAT 6235 6295
CC FT REPEAT 6295 6355
CC FT REPEAT 6355 6415
CC FT REPEAT 6415 6475
CC FT REPEAT 6475 6535
CC FT REPEAT 6535 6595
CC FT REPEAT 6595 6655
CC FT REPEAT 6655 6715
CC FT REPEAT 6715 6775
CC FT REPEAT 6775 6835
CC FT REPEAT 6835 6895
CC FT REPEAT 6895 6955
CC FT REPEAT 6955 7015
CC FT REPEAT 7015 7075
CC FT REPEAT 7075 7135
CC FT REPEAT 7135 7195
CC FT REPEAT 7195 7255
CC FT REPEAT 7255 7315
CC FT REPEAT 7315 7375
CC FT REPEAT 7375 7435
CC FT REPEAT 7435 7495
CC FT REPEAT 7495 7555
CC FT REPEAT 7555 7615
CC FT REPEAT 7615 7675
CC FT REPEAT 7675 7735
CC FT REPEAT 7735 7795
CC FT REPEAT 7795 7855
CC FT REPEAT 7855 7915
CC FT REPEAT 7915 7975
CC FT REPEAT 7975 8035
CC FT REPEAT 8035 8095
CC FT REPEAT 8095 8155
CC FT REPEAT 8155 8215
CC FT REPEAT 8215 8275
CC FT REPEAT 8275 8335
CC FT REPEAT 8335 8395
CC FT REPEAT 8395 8455
CC FT REPEAT 8455 8515
CC FT REPEAT 8515 8575
CC FT REPEAT 8575 8635
CC FT REPEAT 8635 8695
CC FT REPEAT 8695 8755
CC FT REPEAT 8755 8815
CC FT REPEAT 8815 8875
CC FT REPEAT 8875 8935
CC FT REPEAT 8935 8995
CC FT REPEAT 8995 9055
CC FT REPEAT 9055 9115
CC FT REPEAT 9115 9175
CC FT REPEAT 9175 9235
CC FT REPEAT 9235 9295
CC FT REPEAT 9295 9355
CC FT REPEAT 9355 9415
CC FT REPEAT 9415 9475
CC FT REPEAT 9475 9535
CC FT REPEAT 9535 9595
CC FT REPEAT 9595 9655
CC FT REPEAT 9655 9715
CC FT REPEAT 9715 9775
CC FT REPEAT 9775 9835
CC FT REPEAT 9835 9895
CC FT REPEAT 9895 9955
CC FT REPEAT 9955 10015
CC FT REPEAT 10015 10075
CC FT REPEAT 10075 10135
CC FT REPEAT 10135 10195
CC FT REPEAT 10195 10255
CC FT REPEAT 10255 10315
CC FT REPEAT 10315 10375
CC FT REPEAT 10375 10435
CC FT REPEAT 10435 10495
CC FT REPEAT 10495 10555
CC FT REPEAT 10555 10615
CC FT REPEAT 10615 10675
CC FT REPEAT 10675 10735
CC FT REPEAT 10735 10795
CC FT REPEAT 10795 10855
CC FT REPEAT 10855 10915
CC FT REPEAT 10915 10975
CC FT REPEAT 10975 11035
CC FT REPEAT 11035 11095
CC FT REPEAT 11095 11155
CC FT REPEAT 11155 11215
CC FT REPEAT 11215 11275
CC FT REPEAT 11275 11335
CC FT REPEAT 11335 11395
CC FT REPEAT 11395 11455
CC FT REPEAT 11455 11515
CC FT REPEAT 11515 11575
CC FT REPEAT 11575 11635
CC FT REPEAT 11635 11695
CC FT REPEAT 11695 11755
CC FT REPEAT 11755 11815
CC FT REPEAT 11815 11875
CC FT REPEAT 11875 11935
CC FT REPEAT 11935 11995
CC FT REPEAT 11995 12055
CC FT REPEAT 12055 12115
CC FT REPEAT 12115 12175
CC FT REPEAT 12175 12235
CC FT REPEAT 12235 12295
CC FT REPEAT 12295 12355
CC FT REPEAT 12355 12415
CC FT REPEAT 12415 12475
CC FT REPEAT 12475 12535
CC FT REPEAT 12535 12595
CC FT REPEAT 12595 12655
CC FT REPEAT 12655 12715
CC FT REPEAT 12715 12775
CC FT REPEAT 12775 12835
CC FT REPEAT 12835 12895
CC FT REPEAT 12895 12955
CC FT REPEAT 12955 13015
CC FT REPEAT 13015 13075
CC FT REPEAT 13075 13135
CC FT REPEAT 13135 13195
CC FT REPEAT 13195 13255
CC FT REPEAT 13255 13315
CC FT REPEAT 13315 13375
CC FT REPEAT 13375 13435
CC FT REPEAT 13435 13495
CC FT REPEAT 13495 13555
CC FT REPEAT 13555 13615
CC FT REPEAT 13615 13675
CC FT REPEAT 13675 13735
CC FT REPEAT 13735 13795
CC FT REPEAT 13795 13855
CC FT REPEAT 13855 13915
CC FT REPEAT 13915 13975
CC FT REPEAT 13975 14035
CC FT REPEAT 14035 14095
CC FT REPEAT 14095 14155
CC FT REPEAT 14155 14215
CC FT REPEAT 14215 14275
CC FT REPEAT 14275 14335
CC FT REPEAT 14335 14395
CC FT REPEAT 14395 14455
CC FT REPEAT 14455 14515
CC FT REPEAT 14515 14575
CC FT REPEAT 14575 14635
CC FT REPEAT 14635 14695
CC FT REPEAT 14695 14755
CC FT REPEAT 14755 14815
CC FT REPEAT 14815 14875
CC FT REPEAT 14875 14935
CC FT REPEAT 14935 14995
CC FT REPEAT 14995 15055
CC FT REPEAT 15055 15115
CC FT REPEAT 15115 15175
CC FT REPEAT 15175 15235
CC FT REPEAT 15235 15295
CC FT REPEAT 15295 15355
CC FT REPEAT 15355 15415
CC FT REPEAT 15415 15475
CC FT REPEAT 15475 15535
CC FT REPEAT 15535 15595
CC FT REPEAT 15595 15655
CC FT REPEAT 15655 15715
CC FT REPEAT 15715 15775
CC FT REPEAT 15775 15835
CC FT REPEAT 15835 15895
CC FT REPEAT 15895 15955
CC FT REPEAT 15955 16015
CC FT REPEAT 16015 16075
CC FT REPEAT 16075 16135
CC FT REPEAT 16135 16195
CC FT REPEAT 16195 16255
CC FT REPEAT 16255 16315
CC FT REPEAT 16315 16375
CC FT REPEAT 16375 16435
CC FT REPEAT 16435 16495
CC FT REPEAT 16495 16555
CC FT REPEAT 16555 16615
CC FT REPEAT 16615 16675
CC FT REPEAT 16675 16735
CC FT REPEAT 16735 16795
CC FT REPEAT 16795 16855
CC FT REPEAT 16855 16915
CC FT REPEAT 16915 16975
CC FT REPEAT 16975 17035
CC FT REPEAT 17035 17095
CC FT REPEAT 17095 17155
CC FT REPEAT 17155 17215
CC FT REPEAT 17215 17275
CC FT REPEAT 17275 17335
CC FT REPEAT 17335 17395
CC FT REPEAT 17395 17455
CC FT REPEAT 17455 17515
CC FT REPEAT 17515 17575
CC FT REPEAT 17575 17635
CC FT REPEAT 17635 17695
CC FT REPEAT 17695 17755
CC FT REPEAT 17755 17815
CC FT REPEAT 17815 17875
CC FT REPEAT 17875 17935
CC FT REPEAT 17935 17995
CC FT REPEAT 17995 18055
CC FT REPEAT 18055 18115
CC FT REPEAT 18115 18175
CC FT REPEAT 18175 18235
CC FT REPEAT 18235 18295
CC FT REPEAT 18295 18355
CC FT REPEAT 18355 18415
CC FT REPEAT 18415 18475
CC FT REPEAT 18475 18535
CC FT REPEAT 18535 18595
CC FT REPEAT 18595 18655
CC FT REPEAT 18655 18715
CC FT REPEAT 18715 18775
CC FT REPEAT 18775 18835
CC FT REPEAT 18835 18895
CC FT REPEAT 18895 18955
CC FT REPEAT 18955 19015
CC FT REPEAT 19015 19075
CC FT REPEAT 19075 19135
CC FT REPEAT 19135 19195
CC FT REPEAT 19195 19255
CC FT REPEAT 19255 19315
CC FT REPEAT 19315 19375
CC FT REPEAT 19375 19435
CC FT REPEAT 19435 19495
CC FT REPEAT 19495 19555
CC FT REPEAT 19555 19615
CC FT REPEAT 19615 19675
CC FT REPEAT 19675 19735
CC FT REPEAT 19735 19795
CC FT REPEAT 19795 19855
CC FT REPEAT 19855 19915
CC FT REPEAT 19915 19975
CC FT REPEAT 19975 20035
CC FT REPEAT 20035 20095
CC FT REPEAT 20095 20155
CC FT REPEAT 20155 20215
CC FT REPEAT 20215 20275
CC FT REPEAT 20275 20335
CC FT REPEAT 20335 20395
CC FT REPEAT 20395 20455
CC FT REPEAT 20455 20515
CC FT REPEAT 20515 20575
CC FT REPEAT 20575 20635
CC FT REPEAT 20635 20695
CC FT REPEAT 20695 20755
CC FT REPEAT 20755 20815
CC FT REPEAT 20815 20875
CC FT REPEAT 20875 20935
CC FT REPEAT 20935 20995
CC FT REPEAT 20995 21055
CC FT REPEAT 21055 21115
CC FT REPEAT 21115 21175
CC FT REPEAT 21175 21235
CC FT REPEAT 21235 21295
CC FT REPEAT 21295 21355
CC FT REPEAT 21355 21415
CC FT REPEAT 21415 21475
CC FT REPEAT 21475 21535
CC FT REPEAT 21535 21595
CC FT REPEAT 21595 21655
CC FT REPEAT 21655 21715
CC FT REPEAT 21715 21775
CC FT REPEAT 21775 21835
CC FT REPEAT 21835 21895
CC FT REPEAT 21895 21955
CC FT REPEAT 21955 22015
CC FT REPEAT 22015 22075
CC FT REPEAT 22075 22135
CC FT REPEAT 22135 22195
CC FT REPEAT 22195 22255
CC FT REPEAT 22255 22315
CC FT REPEAT 22315 22375
CC FT REPEAT 22375 22435
CC FT REPEAT 22435 22495
CC FT REPEAT 22495 22555
CC FT REPEAT 22555 22615
CC FT REPEAT 22615 22675
CC FT REPEAT 22675 22735
CC FT REPEAT 22735 22795
CC FT REPEAT 22795 22855
CC FT REPEAT 22855 22915
CC FT REPEAT 22915 22975
CC FT REPEAT 22975 23035
CC FT REPEAT 23035 23095
CC FT REPEAT 23095 23155
CC FT REPEAT 23155 23215
CC FT REPEAT 23215 23275
CC FT REPEAT 23275 23335
CC FT REPEAT 23335 23395
CC FT REPEAT 23395 23455
CC FT REPEAT 23455 23515
CC FT REPEAT 23515 23575
CC FT REPEAT 23575 23635
CC FT REPEAT 23635 23695
CC FT REPEAT 23695 23755
CC FT REPEAT 23755 23815
CC FT REPEAT 23815 23875
CC FT REPEAT 23875 23935
CC FT REPEAT 23935 23995
CC FT REPEAT 23995 24055
CC FT REPEAT 24055 24115
CC FT REPEAT 24115 24175
CC FT REPEAT 24175 24235
CC FT REPEAT 24235 24295
CC FT REPEAT 24295 24355
CC FT REPEAT 24355 24415
CC FT REPEAT 24415 24475
CC FT REPEAT 24475 24535
CC FT REPEAT 24535 24595
CC FT REPEAT 24595 24655
CC FT REPEAT 24655 24715
CC FT REPEAT 24715 24775
CC FT REPEAT 24775 24835
CC FT REPEAT 24835 24895
CC FT REPEAT 24895 24955
CC FT REPEAT 24955 25015
CC FT REPEAT 25015 25075
CC FT REPEAT 25075 25135
CC FT REPEAT 25135 25195
CC FT REPEAT 25195 25255
CC FT REPEAT 25255 25315
CC FT REPEAT 25315 25375
CC FT REPEAT 25375 25435
CC FT REPEAT 25435 25495
CC FT REPEAT 25495 25555
CC FT REPEAT 25555 25615
CC FT REPEAT 25615 25675
CC FT REPEAT 25675 25735
CC FT REPEAT 25735 25795
CC FT REPEAT 25795 25855
CC FT REPEAT 25855 25915
CC FT REPEAT 25915 25975
CC FT REPEAT 25975 26035
CC FT REPEAT 26035 26095
CC FT REPEAT 26095 26155
CC FT REPEAT 26155 26215
CC FT REPEAT 26215 26275
CC FT REPEAT 26275 26335
CC FT REPEAT 26335 26395
CC FT REPEAT 26395 26455
CC FT REPEAT 26455 26515
CC FT REPEAT 26515 26575
CC FT REPEAT 26575 26635
CC FT REPEAT 26635 26695
CC FT REPEAT 26695 26755
CC FT REPEAT 26755 26815
CC FT REPEAT 26815 26875
CC FT REPEAT 26875 26935
CC FT REPEAT 26935 26995
CC FT REPEAT 26995 27055
CC FT REPEAT 27055 27115
CC FT REPEAT 27115 27175
CC FT REPEAT 27175 27235
CC FT REPEAT 27235 27295
CC FT REPEAT 27295 27355
CC FT REPEAT 27355 27415
CC FT REPEAT 27415 27475
CC FT REPEAT 27475 27535
CC FT REPEAT 27535 27595
CC FT REPEAT 27595 27655
CC FT REPEAT 27655 27715
CC FT REPEAT 27715 27775
CC FT REPEAT 27775 27835
CC FT REPEAT 27835 27895
CC FT REPEAT 27895 27955
CC FT REPEAT 27955 28015
CC FT REPEAT 28015 28075
CC FT REPEAT 28075 28135
CC FT REPEAT 28135 28195
CC FT REPEAT 28195 28255
CC FT REPEAT 28255 28315
CC FT REPEAT 28315 28375
CC FT REPEAT 28375 28435
CC FT REPEAT 28435 28495
CC FT REPEAT 28495 28555
CC FT REPEAT 28555 28615
CC FT REPEAT 28615 28675
CC FT REPEAT 28675 28735
CC FT REPEAT 28735 28795
CC FT REPEAT 28795 28855
CC FT REPEAT 28855 28915
CC FT REPEAT 28915 28975
CC FT REPEAT 28975 29035
CC FT REPEAT 29035 29095
CC FT REPEAT 29095 29155
CC FT REPEAT 29155 29215
CC FT REPEAT 29215 29275
CC FT REPEAT 29275 29335
CC FT REPEAT 29335 29395
CC FT REPEAT 29395 29455
CC FT REPEAT 29455 29515
CC FT REPEAT 29515 29575
CC FT REPEAT 29575 29635
CC FT REPEAT 29635 29695
CC FT REPEAT 29695 29755
CC FT REPEAT 29755 29815
CC FT REPEAT 29815 29875
CC FT REPEAT 29875 29935
CC FT REPEAT 29935 29995
CC FT REPEAT 29995 30055
CC FT REPEAT 30055 30115
CC FT REPEAT 30115 30175
CC FT REPEAT 30175 30235
CC FT REPEAT 30235 30295
CC FT REPEAT 30295 30355
CC FT REPEAT 30355 30415
CC FT REPEAT 30415 30475
CC FT REPEAT 30475 30535
CC FT REPEAT 30535 30595
CC FT REPEAT 30595 30655
CC FT REPEAT 30655 30715
CC FT REPEAT 30715 30775
CC FT REPEAT 30775 30835
CC FT REPEAT 30835 30895
CC FT REPEAT 30895 30955
CC FT REPEAT 30955 31015
CC FT REPEAT 31015 31075
CC FT REPEAT 31075 31135
CC FT REPEAT 31135 31195
CC FT REPEAT 31195 31255
CC FT REPEAT 31255 31315
CC FT REPEAT 31315 31375
CC FT REPEAT 31375 31435
CC FT REPEAT 31435 31495
CC FT REPEAT 31495 31555
CC FT REPEAT 31555 31615
CC FT REPEAT 31615 31675
CC FT REPEAT 31675 31735
CC FT REPEAT 31735 31795
CC FT REPEAT 31795 31855
CC FT REPEAT 31855 31915
CC FT REPEAT 31915 31975
CC FT REPEAT 31975 32035
CC FT REPEAT 32035 32095
CC FT REPEAT 32095 32155
CC FT REPEAT 32155 32215
CC FT REPEAT 32215 32275
CC FT REPEAT 32275 32335
CC FT REPEAT 32335 32395
CC FT REPEAT 32395 32455
CC FT REPEAT 32455 32515
CC FT REPEAT 32515 32575
CC FT REPEAT 32575 32635
CC FT REPEAT 32635 32695
CC FT REPEAT 32695 32755
CC FT REPEAT 32755 32815
CC FT REPEAT 32815 32875
CC FT REPEAT 32875 32935
CC FT REPEAT 32935 32995
CC FT REPEAT 32995 33055
CC FT REPEAT 33055 33115
CC FT REPEAT 33115 33175
CC FT REPEAT 33175 33235
CC FT REPEAT 33235 33295
CC FT REPEAT 33295 33355
CC FT REPEAT 33355 33415
CC FT REPEAT 33415 33475
CC FT REPEAT 33475 33535
CC FT REPEAT 33535 33595
CC FT REPEAT 33595 33655
CC FT REPEAT 33655 33715
CC FT REPEAT 33715 33775
CC FT REPEAT 33775 33835
CC FT REPEAT 33835 33895
CC FT REPEAT 33895 33955
CC FT REPEAT 33955 34015
CC FT REPEAT 34015 34075
CC FT REPEAT 34075 34135
CC FT REPEAT 34135 34195
CC FT REPEAT 34195 34255
CC FT REPEAT 34255 34315
CC FT REPEAT 34315 34375
CC FT REPEAT 34375 34435
CC FT REPEAT 34435 34495
CC FT REPEAT 34495 34555
CC FT REPEAT 34555 34615
CC FT REPEAT 34615 34675
CC FT REPEAT 34675 34735
CC FT REPEAT 34735 34795
CC FT REPEAT 34795 34855
CC FT REPEAT 34855 34915
CC FT REPEAT 34915 34975
CC FT REPEAT 34975 35035
CC FT REPEAT 35035 35095
CC FT REPEAT 35095 35155
CC FT REPEAT 35155 35215
CC FT REPEAT 35215 35275
CC FT REPEAT 35275 35335
CC FT REPEAT 35335 35395
CC FT REPEAT 35395 35455
CC FT REPEAT 35455 35515
CC FT REPEAT 35515 35575
CC FT REPEAT 35575 35635
CC FT REPEAT 35635 35695
CC FT REPEAT 35695 35755
CC FT REPEAT 35755 35815
CC FT REPEAT 35815 35875
CC FT REPEAT 35875 35935
CC FT REPEAT 35935 35995
CC FT REPEAT 35995 36055
CC FT REPEAT 36055 36115
CC FT REPEAT 36115 36175
CC FT REPEAT 36175 36235
CC FT REPEAT 36235 36295
CC FT REPEAT 36295 36355
CC FT REPEAT 36355 36415
CC FT REPEAT 36415 36475
CC FT REPEAT 36475 36535
CC FT REPEAT 36535 36595
CC FT REPEAT 36595 36655
CC FT REPEAT 36655 36715
CC FT REPEAT 36715 36775
CC FT REPEAT 36775 36835
CC FT REPEAT 36835 36895
CC FT REPEAT 36895 36955
CC FT REPEAT 36955 37015
CC FT REPEAT 37015 37075
CC FT REPEAT 37075 37135
CC FT REPEAT 37135 37195
CC FT REPEAT 37195 37255
CC FT REPEAT 37255 37315
CC FT REPEAT 37315 37375
CC FT REPEAT 37375 37435
CC FT REPEAT 37435 37495
CC FT REPEAT 37495 37555
CC FT REPEAT 37555 37615
CC FT REPEAT 37615 37675
CC FT REPEAT 37675 37735
CC FT REPEAT 37735 37795
CC FT REPEAT 37795 37855
CC FT REPEAT 37855 37915
CC FT REPEAT 37915 37975
CC FT REPEAT 37975 38035
CC FT REPEAT 38035 38095
CC FT REPEAT 38095 38155
CC FT REPEAT 38155 38215
CC FT REPEAT 38215 38275
CC FT REPEAT 38275 38335
CC FT REPEAT 38335 38395
CC FT REPEAT 38395 38455
CC FT REPEAT 38455 38515
CC FT REPEAT 38515 38575
CC FT REPEAT 38575 38635
CC FT REPEAT 38635 38695
CC FT REPEAT 38695 38755
CC FT REPEAT 38755 38815
CC FT REPEAT 38815 38875
CC FT REPEAT 38875 38935
CC FT REPEAT 38935 38995
CC FT REPEAT 38995 39055
CC FT REPEAT 39055 39115
CC FT REPEAT 39115 39175
CC FT REPEAT 39175 39235
CC FT REPEAT 39235 39295
CC FT REPEAT 39295 39355
CC FT REPEAT 39355 39415
CC FT REPEAT 39415 39475
CC FT REPEAT 39475 39535
CC FT REPEAT 39535 39595
CC FT REPEAT 39595 39655
CC FT REPEAT 39655 39715
CC FT REPEAT 39715 39775
CC FT REPEAT 39775 39835
CC FT REPEAT 39835 39895
CC FT REPEAT 39895 39955
CC FT REPEAT 39955 40015
CC FT REPEAT 40015 40075
CC FT REPEAT 40075 40135
CC FT REPEAT 40135 40195
CC FT REPEAT 40195 40255
CC FT REPEAT 40255 40315
CC FT REPEAT 40315 40375
CC FT REPEAT 40375 40435
CC FT REPEAT 40435 40495
CC FT REPEAT 40495 40555
CC FT REPEAT 40555 40615
CC FT REPEAT 40615 40675
CC FT REPEAT 40675 40735
CC FT REPEAT 40735 40795
CC FT REPEAT 40795 40855
CC FT REPEAT 40855 40915
CC FT REPEAT 40915 40975
CC FT REPEAT 40975 41035
CC FT REPEAT 41035 41095
CC FT REPEAT 41095 41155
CC FT REPEAT 41155 41215
CC FT REPEAT 41215 41275
CC FT REPEAT 41275 41335
CC FT REPEAT 41335 41395
CC FT REPEAT 41395 41455
CC FT REPEAT 41455 41515
CC FT REPEAT 41515 41575
CC FT REPEAT 41575 41635
CC FT REPEAT 41635 41695
CC FT REPEAT 41695 41755
CC FT REPEAT 41755 41815
CC FT REPEAT 41815 41875
CC FT REPEAT 41875 41935
CC FT REPEAT 41935 41995
CC FT REPEAT 41995 42055
CC FT REPEAT 42055 42115
CC FT REPEAT 42115 42175
CC FT REPEAT 42175 42235
CC FT REPEAT 422
```

FT REPEAT 327 372 KETCH 1.
 FT REPEAT 373 423 KETCH 2.
 FT REPEAT 424 470 KETCH 3.
 FT REPEAT 471 517 KETCH 4.
 FT REPEAT 518 564 KETCH 5.
 FT REPEAT 565 611 KETCH 6.
 SQ SEQUENCE 624 AA; 69665 MW; 70528F3BF43B6C90 CRC64;

Query Match 14.9%; Score 492; DB 1; Length 624;
 Best Local Similarity 24.7%; Pred. No. 1,7e-28;
 Matches 149; Conservative 112; Mismatches 227; Indels 116; Gaps 20;

10 NTEYAVSLLEQLKLFYEQQLFTDILVLE-----GTEFPCXKXVLAFCSSYFRAMFMSGL 64
 59 HTKQAFGIMNELRL--SQQL-CDVTLQVKYQDAPAFQAFMHKVLVLAASSVFRAMFMTNGL 115
 65 SESEKQTHVLRANDATLQIITVATYAGNLAMDSYEQLYEPAFLQVDEVLQRCREYL 124
 116 REGGMEVVSIEGHPKMERLIEFAAYTASISMEKCVLHMNGAVMYQIDSVRAQSDFL 175
 125 IKKINAMNCVRLSPADLFSCBELKQSAKRMVEHKFTAVYHODAFMQLSHDLIDILSSD 184
 176 VOQLDPSNAGINFAEQICVELHQARERYIMHGEVAKQSEFFVLSHCQLVTLISRD 235
 185 NLNVEKEETVREAAMLMLEYNTESRSQYLSVLSQIRIDALSEVTORAMFQGLPPN--- 240
 236 DLNVRCESEVFHACIMVWKYDCQRRFYVALLRVAVRCHSLT-----PVPFLQW 283
 241 --DKSVVVGGLYKSMKFFPRILGTMKEEMITEASSENPGS-----LYSSVCYSPQ 291
 284 QLOKCEILQSDSRG---KDYLVKIFEEELTLH-KPQVWPCRAPKVGRILYTAGGYFRQ 337
 292 AEKVKYKLCSP-----ADLH-----KVGTVTPPNNDIYIAGQVPLKATKTNHSTKSK 339
 338 SLSTYLEANVPNSCTWRLADLVPRSGLAGCV--GGLYAVAGSR---NNSPQNTDSSA 392
 340 LQTAFTVTCFYFDAQONTWPKPTMPLFVRIKPSLVCCGTYIAGDSVGGELNRRVY 399
 393 L-----DC--YNPMTNQSPCAPMSVPRNRIGVVIDGHIYAVAGSH--GCIHNSV 440
 400 ERYDTEKDEKWTMSPLPCAMQMSAAVYVHDCIYV-----TLNLMYCFPPSDSWEN 452
 441 ERYBRPRDEWHLVAPMLTRRIGVAVLNLRLYAVGGFDGTNRLNAGCYPPERNEMRM 500
 453 AMKQTSRSPASAAFGDKIFYIGL-----HIATNSGIR 486
 501 TANNTIRSGAGVCLNCTIYAGGYGQDQLNVERDYDETETWTVAAPKHKRSALGIT 560
 487 LPSGTV-----DGSV--TVELIYVNMKEMKXANIPAKRYS-----DPCVRAVVIS 531
 561 VHQGRIVYLGVDGHTFLDSVECCYDPTDTWSEVTMTSGRSGVAVTMEPCRKQIDQ 620
 532 NSLC 535
 621 NCTC 624

DB 621 NCTC 624

RESULT 14
 YDS4_HUMAN STANDARD; PRT; 632 AA.
 AC Q9P2J3;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein KIAA1354 (Fragment).
 GN KIAA1354.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxId=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;

RX MEDLINE=20181126; PubMed=10718198;
 RA Nagase T., Kikuno R., Ishikawa K.-I., Hirose M., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XVI.
 RT The complete sequence of 150 new cDNA clones from brain which code
 RT for large proteins in vitro."
 RL DNA Res. 7:65-73(2000).
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 6 KETCH REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-stb.ch/announce/>
 CC or send an email to license@isb-stb.ch).

CC EMBL: AB037775; BAA92592.1; -
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR001798; Ketch.
 DR Pfam; PF00651; BTB; 1.
 DR Pfam; PF01344; Ketch; 6.
 DR SMART; SM00225; BTB; 1.
 DR PROSITE; PS50097; BTB; 1.
 DR Hypothetical protein; Repeat.

FT NON TER 1 1
 FT DOMAIN 1 134 BTB.
 FT REPEAT 315 362 KETCH 1.
 FT REPEAT 363 414 KETCH 2.
 FT REPEAT 415 461 KETCH 3.
 FT REPEAT 463 508 KETCH 4.
 FT REPEAT 510 560 KETCH 5.
 FT REPEAT 561 610 KETCH 6.
 SQ SEQUENCE 632 AA; 71146 MW; EDBE7003432FD73E CRC64;

Query Match 14.4%; Score 474; DB 1; Length 632;
 Best Local Similarity 25.6%; Pred. No. 3.5e-27;
 Matches 163; Conservative 105; Mismatches 259; Indels 110; Gaps 22;

10 NTEYAVSL--LEQLKLFYEQQLFTDILVLE--GTEFPCXKXVLAFCSSYFRAMFMSGL 65
 45 NTHSSVVLGGFQOLRL--EGLCDVTLVPGGDELFPVHRMAMASASYFRAMFGKMK 101
 66 ESKQTHVLRANDATLQIITVATYAGNLAMDSYEQLYEPAFLQVDEVLQRCREYL 125
 102 EDDLMCKLHGVNKKGLKIIDPITAKLSLMDNLDTLEASFLQLPVLDFCVFI 161
 126 KIKINAMNCVRLSPADLFSCBELKQSAKRMVEHKFTAVYHODAFMQLSHDLIDILSSD 185
 162 SCVSLDNCVEGRINTYNTLIEVDKYVNNFLLKNPPALSTGEFLKLPFERLAFLVLSNS 221
 186 LNVEKEETVREAAMLMLEYNTESRSQYLSVLSQIRIDALSEVTORAMFQGLPPNDSVV 245
 222 LKHCTELELFKACRWLLE-DPRMDYAKMKNIR-----FPLMPODLIN 268
 246 VOGL-----YKMPKFKPRL-----GMTKEEMK 269
 269 VQTVDFMRRTDNCVNLLEASNYQMP--YMGVPMQSDRAIRSDTHLVTLGGVLRQQLV 327
 270 ITEASSENPCSLYSVCYSPOAEKVKYKLCSPADLHKVGVVTPNDIYIAGGVPLKN 329
 328 VKEELR-----MYDERAQEWRSLAPDADRYQGIYVI-GNFIYVVGQ----- 370
 330 TITNNSKTSKQOTARTYVNCFWPDAQONTWPKPTMPLFVRIRKPSLVCCGTYIAGDS 389
 371 --SNYD--TKGTYAVDTV---FRDPKRYNMQVAVSLNEKRFFPHLSALKGLHYAAGGS 423
 390 VGGELNRRVREYDTEKDEKWTMSPLPCAMQMSAAVYVHDCIYV-----TLNLMYCY 442
 424 AAGEL--ATVECYNRRMNSYVAKMSBPHYAGLVYISGITHDFQDELKMF 481
 443 PPRSDSWEMAMKQTSRSPASAAFGDKIFYIGLHIAITNSGIRLPSGTVDSSV-TVEI 501

```

Db      482 DEDTDKMKQKAPMTTVRLGLHCMTGVGDKLYVIGGNHF-----RGTSYDDVLSCEY 532
Qy      502 YDVKIKEMWMAANIPAKRYSDDPCVRAVVISNSLCVPMRETHINERAKYTVYOYDELDLDRM 561
      533 YSPITLDQWTFPIIAM-LRGSD--VGVAVPEKNTIYVGVGSMNRCNVEIVQKIDPEKDM 589
Qy      562 SLRQHSERVLWDLGDFRCYVQKLYPSCLSSPMKP 598
      590 HKVFLDPE---SLGGIRACTITLVFPF---EENPSPF 619

RESULT 15
KEAP_MOUSE
ID      KEAP_MOUSE      STANDARD;      PRT;      624 AA.
AC      Q922X8;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      kelch-like ECH-associated protein 1 (Cytosolic inhibitor of Nrf2).
GN      KEAP1.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
RN      11
RP      SEQUENCE FROM N.A.
RC      TISSUE=Embryo;
RX      MEDLINE=99104163; PubMed=9887101;
RA      Itoh K., Wakabayashi N., Katoh Y., Ishii T., Igarashi K., Engel J.D.,
RA      Yamamoto M.;
RT      "Keap1 represses nuclear activation of antioxidant responsive elements
RL      by Nrf2 through binding to the amino-terminal Neh2 domain.";
RN      12
RP      SEQUENCE FROM N.A.
RC      STRAIN=C57BL/60; TISSUE=Lung;
RX      MEDLINE=21085660; PubMed=11217851;
RA      Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA      Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA      Aizawa K., Izawa M., Niehi K., Kiyosawa H., Kondo S., Yamataka I.,
RA      Saito T., Okazaki Y., Gojobori T., Bono H., Kasubawa T., Saito R.,
RA      Kadoya K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA      Fleschmann W., Gaesteland T., Glasl C., King B., Kochiwa H.,
RA      Kuell P., Lewis S., Matsuo Y., Nikiado I., Pasole G., Quackenbush J.,
RA      Schriml L.M., Staahl P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA      Sakai K., Okido T., Furuno M., Aono H., Balderelli R., Barch G.,
RA      Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA      Brownstein M.D., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA      Gueffincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA      Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA      Nordone P., Ring B., Ringwald C., Seta T., Shibata Y., Storch K.-F.,
RA      Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA      Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA      Hayashizaki Y.;
RT      "Functional annotation of a full-length mouse cDNA collection.";
RL      Nature 409:685-690(2001).
CC      -1- FUNCTION: RETAINS NFE2L2/NRF2 IN THE CYTOSOL THUS RESULTING IN THE
CC      SUPPRESSION OF ITS TRANSCRIPTIONAL ACTIVITY AND THE REPRESSION OF
CC      ANTIOXIDANT RESPONSE ELEMENT-MEDIATED DETOXYIFYING ENZYME GENE
CC      EXPRESSION.
CC      -1- SUBUNIT: INTERACTS WITH THE N-TERMINAL REGULATORY DOMAIN OF
CC      NFE2L2/NRF2.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC      -1- SIMILARITY: CONTAINS 6 KELCH REPEATS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/

```

```

CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: AB020063; BAA34639.1; -.
CC      DR      EMBL: AK004738; BAB33519.1; -.
CC      DR      MGD: MGI:1858732; Keap1.
CC      DR      InterPro: IPR000210; BTB_POZ.
CC      DR      InterPro: IPR001798; Kelch.
CC      DR      Pfam: PF00651; BTB; 1.
CC      DR      Pfam: PF01344; Kelch; 6.
CC      DR      PRINTS: PR00501; KELCHREPEAT.
CC      DR      SMART: SM00225; BTB; 1.
CC      DR      PROSITE: PS50097; BTB; 1.
CC      KW      Transcription regulation; Repeat.
CC      FT      DOMAIN 77 149
CC      FT      REPEAT 327 372
CC      FT      REPEAT 373 423
CC      FT      REPEAT 424 470
CC      FT      REPEAT 471 517
CC      FT      REPEAT 519 564
CC      FT      REPEAT 565 611
CC      FT      REPEAT 612 655
CC      SQ      SEQUENCE 624 AA; 69552 MW; 4645DB0122F85F54 CRC64;

Query Match      14.4%; Score 473; DB 1; Length 624;
Best Local Similarity 23.7%; Pred. No. 4.1e-27;
Matches 145; Conservative 111; Mismatches 223; Indels 134; Gaps 19;

Qy      10 NTEVAVSLLEOLKLFYEQQLFTDIYVE-----GTEPCHKMVLATGCSYFRAMFMGL 64
      59 HTKQAFGVWNEDELRL--SQOL-CDVTLQVYKEDIPAAQMAHKVLAASSPVFKAFMTGL 115
Qy      65 SESKOTVHLNRVDAATLQIIITVAVTGNLMAANDSTVEQLYETACFLQVEDYLQRCREYL 124
      116 REQGEVAVSIGIHFKWNERLIEPAVTASIGVEKCVLHWNQAVMYQIDSVRAACDPL 175
Db      125 IKKINAEKCVLLSPADLFSCEBELKQSAKRWVHEKFTVYHODAMQSHLLDILISD 184
      176 VOOLDPSNAIGANFAEOIGCTELHQRAREYIYMHFGVAQBEFFNLSHCQLATLISRD 235
Qy      185 NINVEKERTVBAALMLVETNSESROYLSVLSQIRIDALSEVYQGFAMFQGLPNDXSV 244
      236 DLANVCESEVFRACTIDWVKIDCPQRRFYQALRLRAVCHALT----- 277
Db      245 VVOGLYKSMPEKFFKRL-----GMTKEEM-IPLEASSENPCS-----L 282
      278 -----PRLQVQLQKCEILDADNCQYIVQIFQELTLHKPTQAVPCGAPKVRGLI 328
Qy      283 YSSVYCSQAQAEKVYLCSP-----ADLH-----KVGTVTPNDIYIAGGVPLKXT 330
      329 YTAGGVFPQSLSYLEAVNPBSGWLRLADLVPPRSLAGCVV--GGLLYAVVGR--NNS 383
Db      331 KTNHSEKTKLQTPFTVACFWPMDAQNTWPKPMLVRIKPSLVCCSGYIYAGGDSV 390
      384 PDGNTDSSAL-----DC-----YNPMTQWSPCASMSVPRNRIGVIGDGHLYAVGSSH- 432
Qy      391 GGEILNRRTVERDYDEKDEMTWSPPLPCAMQMSAAVVVHDCIYV-----TLNIMCYF 443
      433 -GCIHSSVEREPRDMDHLVAPMLTRIGVAVLRLLYAVAGVPGFTRLNLSAEQY 491
Qy      444 PRSDSWEMAEKQSRSPASAAAFQDKIFYIGL----- 477
      492 PERNEMRMITPMNTRISGAGVCLNLCIYAAAGGQDQLNSVERVYDVEETETWEVAPWR 551
Db      478 HIATNSGIRLPSGV-----DGSSV--TVRIYDNKEMWMAANIPAKRYS-----D 522
      552 HHRSLGITTVOGKLYVIGYDGHFTLDSVECDPDSPTWSEVTRMTSGRSGVAVTWME 611
Qy      523 PCVRAVVISNSLC 535
      612 PERKQIDQNCNC 624
Db

```

Search completed: July 14, 2003, 18:16:04
Job time : 9.47994 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:07:56 ; Search time 28.9026 Seconds
(without alignments)
4441.379 Million cell updates/sec

Title: US-09-815-379-2

Perfect score: 3292
Sequence: 1 MSTDQERQINTEYAVSLAEQ.....STDGTEEFELDEWALPEV 623

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL_21:*

1: sp_archaea:*

2: sp_bacteria:*

3: sp_fungi:*

4: sp_human:*

5: sp_invertebrate:*

6: sp_mammal:*

7: sp_mhc:*

8: sp_organelle:*

9: sp_phase:*

10: sp_plant:*

11: sp_rodent:*

12: sp_virus:*

13: sp_vertebrate:*

14: sp_unclassified:*

15: sp_rv1rus:*

16: sp_bacteriap:*

17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2732	82.7	511	Q9P239	Q9P239 homo sapien
2	2480	75.6	469	Q9R1W9	Q9R1W9 mus musculu
3	1886	57.3	348	Q9UFW7	Q9UFW7 homo sapien
4	1670	50.7	335	Q9Y382	Q9Y382 homo sapien
5	1664.5	50.6	351	Q9P097	Q9P097 homo sapien
6	778.5	23.6	526	Q9GJ15	Q9GJ15 homo sapien
7	638.5	19.4	684	Q8WVZ9	Q8WVZ9 homo sapien
8	631.5	19.2	674	Q96MP6	Q96MP6 homo sapien
9	630.5	19.2	684	Q9H0I6	Q9H0I6 homo sapien
10	578.5	17.6	623	Q9NGX7	Q9NGX7 drosophila
11	578.5	17.6	623	Q9VUUS	Q9VUUS drosophila
12	565.5	17.2	591	Q8VCK5	Q8VCK5 mus musculu
13	537	16.3	593	Q8TBH5	Q8TBH5 homo sapien
14	526.5	16.0	586	Q9BQF8	Q9BQF8 homo sapien
15	526.5	16.0	744	Q9VENS	Q9VENS drosophila
16	524	15.9	748	Q8TBJ7	Q8TBJ7 homo sapien

17	523	15.9	582	4	Q96RF4	Q96RF4 homo sapien
18	517.5	15.7	564	4	Q96MV2	Q96MV2 homo sapien
19	513.5	15.6	589	4	Q96L69	Q96L69 homo sapien
20	512.5	15.6	584	4	Q96FT8	Q96FT8 homo sapien
21	509.5	15.5	586	11	Q9CZP4	Q9CZP4 mus musculu
22	504.5	15.3	707	4	Q96CT2	Q96CT2 homo sapien
23	500	15.2	571	4	Q9HBX5	Q9HBX5 homo sapien
24	499	15.2	571	11	Q9CR40	Q9CR40 mus musculu
25	498.5	15.1	568	11	Q9R2H4	Q9R2H4 rattus norv
26	498	15.1	625	5	Q9S359	Q9S359 drosophila
27	498	15.1	654	5	Q9S2Y0	Q9S2Y0 drosophila
28	497.5	15.1	531	5	Q61795	Q61795 caenorhabd
29	492	14.9	624	4	Q9BPY9	Q9BPY9 homo sapien
30	484.5	14.7	627	5	Q9VK21	Q9VK21 drosophila
31	480.5	14.6	575	5	Q9VGE5	Q9VGE5 drosophila
32	474	14.4	617	4	Q8TC02	Q8TC02 homo sapien
33	473	14.4	608	11	Q9DA07	Q9DA07 mus musculu
34	473	14.4	608	11	Q9DSV2	Q9DSV2 mus musculu
35	472.5	14.4	614	4	Q96MCO	Q96MCO homo sapien
36	470.5	14.3	610	4	Q8WZ60	Q8WZ60 homo sapien
37	467.5	14.2	589	11	Q8R2P1	Q8R2P1 mus musculu
38	466.5	14.2	589	4	Q9H0H3	Q9H0H3 homo sapien
39	460.5	14.0	538	5	Q9VGE6	Q9VGE6 drosophila
40	460	14.0	621	11	Q9D783	Q9D783 mus musculu
41	455	13.8	604	4	Q96Q17	Q96Q17 homo sapien
42	453	13.8	634	11	Q99JN2	Q99JN2 mus musculu
43	452	13.7	634	4	Q9H511	Q9H511 homo sapien
44	451	13.7	613	11	Q8R124	Q8R124 mus musculu
45	448	13.6	638	11	Q9DBY7	Q9DBY7 mus musculu

ALIGNMENTS

RESULT 1

Q9P239	PRELIMINARY;	PRT;	511 AA.
AC	Q9P239		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	KIAA1489 protein (Fragment).		
GN	KIAA1489.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20277482; PubMed=10819331;		
RA	Nagase T., Kikuno R., Ishikawa K., Hirogawa M., Ohara O.;		
RT	"Prediction of the coding sequences of unidentified human		
RT	genes.XVII: The complete sequences of 100 new cDNA clones from brain		
RT	which code for large proteins in vitro."		
RL	DNA Res. 7:143-150(2000).		
DR	EMBL; AB040922; BAA96013.1; -		
DR	InterPro; IPR001798; Kelch.		
DR	Pfam; PF01344; Kelch; 3.		
FT	NON_TER		
FT	NON_TER		
SQ	SEQUENCE	511 AA; 58516 MW; 96E37D756D242DBA CRC64;	

Query Match 82.7%; Score 2732; DB 4; Length 511;

Best Local Similarity 100.0%; Pred. No. 1,1e-221; Indels 0; Gaps 0;

Matches 511; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	113	VEDVLRQREYLIKINENCVRLISFADLFSCEELKQSAKRMVHKETAVYHODAFMQL	172
DB	1	VEDVLRQREYLIKINENCVRLISFADLFSCEELKQSAKRMVHKETAVYHODAFMQL	60
QY	173	SHDLDLIDLSNDLVNVEKEETVREANMLLENTSRSQYSSVLSQIRIDALSTVTOA	232
DB	61	SHDLDLIDLSNDLVNVEKEETVREANMLLENTSRSQYSSVLSQIRIDALSTVTOA	120

Qy	233	MFQGLPNDKSAVYVOGLYSMPKFPKRLGMTKEEMWIFLWASSNPKSLYSSVCVSPQA	292
Db	121	MFQGLPNDKSAVYVOGLYSMPKFPKRLGMTKEEMWIFLWASSNPKSLYSSVCVSPQA	180
Qy	293	EKVYKLCSPADLHKGVTVTPDNDIYIAGGVPLNKTITNSKTSKLTQAPRTVACFYW	352
Db	181	EKVYKLCSPADLHKGVTVTPDNDIYIAGGVPLNKTITNSKTSKLTQAPRTVACFYW	240
Qy	353	FDAQONTWPKTKPMLFVRIKPSLVCCEGYIYAIAGDSVGGELNRRVVERDYDEKEMTWY	412
Db	241	FDAQONTWPKTKPMLFVRIKPSLVCCEGYIYAIAGDSVGGELNRRVVERDYDEKEMTWY	300
Qy	413	SPLPACAMQMSAAVVVHDCIYMTLNLNMYCFPPSDSDVWEMAMQTSRSPFAAAGDKTF	472
Db	301	SPLPACAMQMSAAVVVHDCIYMTLNLNMYCFPPSDSDVWEMAMQTSRSPFAAAGDKTF	360
Qy	473	YIGGLHIAITNSGIRLPSGTVDSSVVEIYDVNKKMKMAAANIIPARYSDPICRAVAIN	532
Db	361	YIGGLHIAITNSGIRLPSGTVDSSVVEIYDVNKKMKMAAANIIPARYSDPICRAVAIN	420
Qy	533	SLCFVRETHLNERAKYVTVQYDLEIDRWLSLQHSERVLMDIGRDPCTVGLYPSCLE	592
Db	421	SLCFVRETHLNERAKYVTVQYDLEIDRWLSLQHSERVLMDIGRDPCTVGLYPSCLE	480
Qy	593	ESPMKRPPTYLFSTDGTGEFELDGEWALPVPV	623
Db	481	ESPMKRPPTYLFSTDGTGEFELDGEWALPVPV	511

RESULT 2
Q8R1W9
ID Q8R1W9 PRELIMINARY; PRT; 469 AA.

DT 01-JUN-2002 (TEMBLrel. 21, Created)
DT 01-JUN-2002 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Hypothetical 53.6 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EYE;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022962; AH02962.1; -
KM Hypothetical protein.
SQ SSOURCE 469 AA; 53590 MW; 94A39E15DBCAB844 CRC64;

Query Match	75.6%	Score 2490	DB 11	Length 469
Best Local Similarity	99.1%	Pred. No. 4.2e-202		

Qy	155	VEVEHEFTVYHODAPMOJSHDILILSSDNLNVEKEEYVREANMTEMENESBOYLS	214
Db	1	MEHEKFTAVYROEAFNOJSHDILILSSDNLNVEKEEYVREANMTEMENESBOYLS	60
Qy	215	SVLSQIRIDALSEVTORAWFOGLPPNDKSVVYGLYKSNPKFEPKPLGTMKEEMFIEA	274
Db	61	SVLSQIRIDALSEVTORAWFOGLPPNDKSVVYGLYKSNPKFEPKPLGTMKEEMFIEA	120
Qy	275	SSENCSLYSVSVCSPOAEKVYKLCSPPADLHKGVNTPPNDIYIAGOVPLKNTKTNH	334
Db	121	SSENCSLYSVSVCSPOAEKVYKLCSPPADLHKGVNTPPNDIYIAGOVPLKNTKTNH	180
Qy	335	SKTSKLQTAFRVNFCEYFMDAQNMTFPKTPLYRIRKESLYCCBEGYIYAIGDSVGGEL	394
Db	181	SKTSKLQTAFRVNFCEYFMDAQNMTFPKTPLYRIRKESLYCCBEGYIYAIGDSVGGEL	240
Qy	395	NRRATYERDTEDEMTNWSPLPCANQMSAAVYVHDCIYMTLNTLWICYPPRSDSWEMAM	454
Db	241	NRRATYERDTEDEMTNWSPLPCANQMSAAVYVHDCIYMTLNTLWICYPPRSDSWEMAM	300

Qy	455	RQSRSEFAAAAFGDKIFPIYGGIHLATNSGRLRPSGTQDGSVVEIYDVKHKEMKAAAN	51.4
Db	301	RQSRSEFAAAAFGDKIFPIYGGIHLATNSGRLRPSGTQDGSVVEIYDVKHKEMKAAAN	36.6
Qy	515	IPAKRSDSCBAVAVVISNLSLVPMETHLNBRACKVYQYDLEIDRMSLRQHSERVAMD	57.4
Db	361	IPAKRSDSCBAVAVVISNLSLVPMETHLNBRACKVYQYDLEIDRMSLRQHSERVAMD	42.0
Qy	575	LGDRPFCYTGKLYPSCLEBSPMKRPPTYFSTDTGTEEFELDSEMAALPPV	62.3
Db	421	LGDRPFCYTGKLYPSCLEBSPMKRPPTYFSPDGTIEEFELDSEMAALPPV	46.9

RESULT 3

ID Q9UFM7 PRELIMINARY; PRT; 348 AA

DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 39.5 kDa protein (Fragment).
GN DKF2P566C134.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
OX [1]
RN RP
RP SEQUENCE FROM N.A.
RC TTSUB=KIDNEY;
RC B10E0CR.H.; Boecher M., Brandt P., Mewes H.W., Gassenhuber J.,
RA Mleamann S.;
RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ
RL EMBL; AL117562; CAB55994.1; -
DR InterPro: IPR001798; Ketch.
DR Pfam: PF01344; Ketch. 3.
DR Hypothetical protein.
FT NCD_NER 1
SQ SSQDNCER 348 AA; 39503 MW; C129A7A6CAE0A1A77 CRC64;

Query Match	57.3%	Score	1886	DB 4	Length	348	
Best Local Similarity	99.4%	Pred. No.	3.5e-151				
Matches 346; Conservative	1	Mismatches	1	Indels	0	Gaps	0

QY	276	SENPCSLYSVCYSPQAEKVYKLCSPPADLHKGTVTTPNDNDIYIAGGOVPLKNTKTNHS	335
:	:	:	:
Db	1	AENPCSLYSVCYSPQAEKVYKLCSPPADLHKGTVTTPNDNDIYIAGGOVPLKNTKTNHS	60
QY	336	KNSKLOTARTNCCYWFDAQONTWFFPRTPLFRIKPSLVCCBGYIYALIGDSVGGELN	395
:	:	:	:
Db	61	KTSKLOTARTNCCYWFDAQONTWFFPRTPLFRIKPSLVCCBGYIYALIGDSVGGELN	120
QY	396	RRTVERYDEKDEMTWSPBPCAMOMSAVVVHDCIYMTLNTMYCFPPSDGSVWEAMR	455
:	:	:	:
Db	121	RRTVERYDEKDEMTWSPBPCAMOMSAVVVHDCIYMTLNTMYCFPPSDGSVWEAMR	180
QY	456	QTSRSPASAAAFGDKIFYIGGLHATNSGIRLPSGTVDGSSVYELIDVANKMEKMAANI	515
:	:	:	:
Db	181	QTSRSPASAAAFGDKIFYIGGLHATNSGIRLPSGTVDGSSVYELIDVANKMEKMAANI	240
QY	516	PAKRYSDPCVRAVVISNLSLVCFMRETHLNERAKVVTQYDLELDRLMSLRHISERVYLMDL	575
:	:	:	:
Db	241	PAKRYSDPCVRAVVISNLSLVCFMRETHSNEAKVYTYQYDLELDRLMSLRHISERVYLMDL	300
QY	576	GNDPFRCTVQKLYPSCLEBSPMKPPTYLFSTDTGTEFELDSEMAVLPV	623
:	:	:	:
Db	301	GNDPFRCTVQKLYPSCLEBSPMKPPTYLFSTDTGTEFELDSEMAVLPV	348
RESULT 4			
QY382			
QY382	PRELIMINARY;	PRT;	335 AA.
QC	QY382;		

DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DS CGI-73 protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:26272150; PubMed:10810093;
RA Lai C.-H.; Chou C.-Y.; Chang U.-Y.; Lin C.-S.; Lin W.-C.;
RT "Identification of Novel Human Genes Evolutionarily Conserved in
RT Caenorhabditis elegans by Comparative Proteomics."
RL Genome Res. 10:703-713(2000).
DR EMBL; AF151831; AAD34068.1; -.
DR InterPro: IPR000210; BTB_POZ.
DR Pfam: PF00651; BTB; 1.
DR SMART: SMO0225; BTB; 1.
DR PROSITE: PS50097; BTB; 1.
QO SEQUENCE 335 AA; 38339 MW; C1B10D6291736A44 CRC64;

```

Db      61 DQQAQKVMISHLDPQNSIGVFIADHYHQGLSDGRSKETIIRKKFLCTKTEQEPFLDGLTKQ 120
Oy      177 LIDIISSDLNANKEKETTREBAMMLEYNTESHESQYLSLVLSO-ITIDALSEVTOQAWFO 235
Db      121 LISIISDDLDLVNDREHVEYESIIRWPEHEONEREVHLPEIFACIRFFPLMED----FTIE 176
Oy      236 GLPNDKSVVVOGLVSMRPFK---PRIGMTKEEMWIFIBASBNPCSLYSVCYSPQA 292
Db      177 KIPQPAQAIKASCVKEGFSNTNGCTQRIQAGMTASEMIFCDPAHKSGKKQVPCLDIYT 236
Oy      293 EKYYKLCSPDADLHKVGTAVTTPNDNDIYIAGGVPLKNTKTNKTSKSLQTAFTVACFYW 352
Db      237 GAVFLCKCPNDLREVGLVSPDNDIYIAGGRP-----SSSEVSIIDHKAEKDFMW 287
Oy      353 PDAQONTPEPKPMLFVRIKPSLVCEGEGYIYAGDSVGE--LNRTRYERYTEKDEWT 410
Db      288 YDHSNTRMLSKPSLLRARIGCKLVLYCCGMVYIIGGVYRQDGRNLSKSECVYSRNCMT 347
Oy      411 MVSPLPCANQWAAVYVNDICVYMTLNTLWVYCFPRSDSVEMAMRQTSFSAFAAFQK 470
Db      348 TVCANPVAHEFNAEYKKEKITYLQGEFFLPFEPQKDYGFPLPMTVPRIQGLAAVYKOS 407
Oy      471 IFYIGGLHATISGRLPESGTVDSSVTEIYDVNKNEMKMANIPAKRYSDPCRAVYI 530
Db      408 IYIYG---TCGNHQRM-----FTYEAVDIELNKTTRKDDPFCDOSINYLYLVLE 455
Oy      531 SNSLCFVMEETHL-----NEBAKYVYQYDLEIDRWSLROHISERVIMLDGRDFCTV 583
Db      456 QNKHLFVATQVTEVEHFRTSRKSLYQYDIDQMKVYETPDR-LMDLGRHFECANV 514
Oy      584 GCLYPSCLER 593
Db      515 AKLYPQCLQK 524

RESULT 7
O8MWZ9 PRELIMINARY; PRT: 684 AA.
AC O8MWZ9;
DT 01-MAR-2002 (TREMBLrel. 20, Created)
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 77.2 kDa protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Struhsberg R.;
RL EMBL; BC022033; AAH22033.1; -.
DR InfePro; IPR000210; BTF_POZ.
DR InfePro; IPR001798; Kelch.
DR Pfam; PF00651; BTF; 1.
DR Pfam; PF01344; Kelch; 3.
DR SMART; SM00225; BTF; 1.
DR PROSITE; PS50097; BTF; 1.
KW Hypothetical protein.
SQ SEQUENCE 684 AA; 77163 MW; B2DBB6969A1AE5E9 CRC64;

Query Match 19.4%; Score 638.5; DB 4; Length 684;
Best Local Similarity 26.1%; Pred. No. 3.4e-45;
Matches 177; Conservative 125; Mismatches 248; Indels 127; Gaps 21

Oy 3 TODERQINTEVAVSLLEQLKLFEEQOLFDTIYLV-----EGTEPFCRKNVATCGSS 54
Db 35 TGPEELKQTAHSAALIAQLKSFYDARLLCDVITEVTPESGPGTGRLLFCNRRVLAACP 94

Oy 55 YFRAMFMGSLSSKQTHNLKRVNDATLQIITTYATGTLNANDSVTEQLYETACQLQYE 114
Db 95 YFKSMFTGGMVESQDASVTMHDVDESEFVADVCTGTGVSSEANVQGLTYASDMLQDE 154

```

Qy	115	DVLCQREYLLIKKINACNVCRLSPADLFCSCGELKQSKMVEKFTPAV-----YHODAF	168
Db	155	YVEACASFLAKRLDITNCTAILKPADAFEDHKKLRSAQSTIYANFQLSRMGSIREETL	214
Qy	170	MOLSHDLLIDILSSDNLNVEKEETREAAIMLIEYNTESSQVLSVLSQIRIDALSEVT	229
Db	215	ADLTLLAQLLAVLRDLSDIDIESERTVCHVAVOMLEAAAKERGPSAAEVFKCVRMHGFEED	274
Qy	230	QRAHFGGL--PRPDSV---VYQG-----LYKSN-----	253
Db	275	Q-DYLEGGLLTKPLVKKYCLVDILEGALQNRYGDLLYKSLVLPVNSSSSSSSNSLVSAAEN	333
Qy	254	-PKFKPRLLGMTKEEMWIFTEASSENPCSLYSVCYSPQAEKYKLCSPADLHKYGTV-	311
Db	334	PPQ-----RLGMCKAEVIFP--GHPRDP-----FLCYDPSGDIYTMSPILTSFAHKTIVT	383
Qy	312	----VTPNDIYIAGGQVPLKNTKTNHSTKTSKLQTAFTVNCYFWPDAQNTWPFKTPM	366
Db	384	SSAVCSPPHDIYLA--QPRKD-----LMVYKPAQNSMOQLADR	421
Qy	367	LPRIRPSLVCCGEGYIYAGG--DSVSGELNRTYERYDTEKDEWTWSPRLPCAMQNSAAV	425
Db	422	LICREGMDAAYIANGYIYILGGRDPTTG--VKKEVECSVQGNQALVAPVPHSFYSFELI	480
Qy	426	VHADCIYVWTLIMCYFPRSPSWYEMMRQTSRSPASAAFGDKIFYIGLHATNSGI	485
Db	481	VQNYLYANXSRMLCYDPSHHMMLNCALSIRS--DFQCAVCYFNDIYICDIPV-----	533
Qy	486	RLPSGTVDSSVTEIYDVNKNKEMWMAANIIPAKIYSDPCYRAVVISNLCVPMRETHLNE	545
Db	534	-----MKYMPAREGEMIRISINILPDSSTHN--YQIVNHQDKLLITSTTPQWK	579
Qy	546	RAKYVYQYVDLDELDRMSLRQHSIEVLMDLGDRDRCYVGKLYPSCLEESWPKPTIYFST	605
Db	580	KNRVLYEYEDTRBDQINIGTMLGLLPDSCG--FLICLCARVPSCLE-----PGOSFIT	631
Qy	606	-----DGTSEFFELDG	615
Db	632	EEDDARSSESTEMDLDG	648
RESULT 8			
Q96MP6		PRELIMINARY;	PRT; 674 AA.
AC	Q96MP6		
DT	01-DEC-2001 (T-EMBLrel. 19, Created)		
DT	01-DEC-2001 (T-EMBLrel. 19, Last sequence update)		
DT	01-JUN-2002 (T-EMBLrel. 21, Last annotation update)		
DE	CDNA FLJ32071 fis, clone OCBBF100122, weakly similar to RING CANAL		
DE	protein.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI	_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN;		
RA	Oshima A., Takahashi-Fujii A., Tanase T., Imose N., Takeuchi K.,		
RA	Arata M., Mitsuhashi K., Yuki H., Hara H., Sugiyama T., Irie R.,		
RA	Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,		
RA	Isono Y., Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K.,		
RA	Yamauchi H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,		
RA	Megatsuma M., Murakawa K., Kanehori K., Sugiyama A., Kawakami B.,		
RA	Suzuki Y., Sugano S., Negahara K., Masuno T., Nagai K., Isogai T.;		
RT	"MEDO human cDNA sequencing project."		
RL	Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.		
RL	EMBL; AK056633; BAB71238.1; -		
DR	InterPro; IPR000210; BTB_POZ.		
DR	InterPro; IPR001798; Kelch.		
DR	Pfam; PF00651; BTB, 1.		
DR	Pfam; PF01344; Kelch, 1.		
DR	PROSITE; PS50097; BTB, 1.		

SQ SEQUENCE 674 AA; 76152 MW; 98275596E094AFB CRC64;
 Query Match 19.2%; Score 631.5; DB 4; Length 674;
 Best Local Similarity 26.1%; Pred. No. 1.3e-44;
 Matches 177; Conservative 128; Mismatches 244; Indels 129; Gaps 22;
 QY 3 TQDERQINTEYAVSLLEQLKLYEQQLFTDIVLV-----EGTEPCHKMYLATCSS 54
 DB 35 TQPEELKDTAHSALLAQKSFYDARQLCDVTIEVTPGSGPTGRLPFCNNRVLAACP 94
 QY 55 YRAMPMSGLSESKQTHVLRNDATLQIITVATGNLANNDSVEQLYETACFLQVE 114
 DB 95 YKSMFTGMYESQOASVTMHVDVDAESFEVLVDYCTGVSLSEANVERLYAASDMLQLE 154
 QY 115 DYLRQCREYLKIKINAENCVRLLSPADLPSCSELKQSAKRWENHKTAAYH-----QDAF 169
 DB 155 YRREACASFLARLDLTNCTAILKFPDAGHRLKRSQOASYAHQNFQSLSHMSIREETL 214
 QY 170 MQLSHDLIDLISDNLVNKEETVEAAMLMLENTESRQOYSSVLSQIRIDALSEVT 229
 DB 215 ADLTIAQLLAVALRLDSDVSEEGTCHVAVOMLEAPKRGPSAAEVFCVMMHPTBED 274
 QY 230 QRAMFGRL--PNDKSV---VVOG-----LYKSM----- 253
 DB 275 Q-DYLEGLTTRKYVKKYCLDVIIEGALQMRVYDLYSLVPVPNSSSSSSNSLVSAEN 333
 QY 254 -PKFPRKLGMTKEEMMIFLEASSENPCSLYSVCYSPQAEKYKLCSPPADLHKVTV 312
 DB 334 PQO---RLGMCCKEMVIFP-GHPRDP-----FLCDDYSGDLYKVPSPFLTCLAHRTVT 383
 QY 313 T-----PNDNIYIAGGOVPLKNTKTNHSTKSLQTAFTVNCFFWPAQONTWPKTPM 366
 DB 384 TLAVCISPDHDIYLA--QPRD-----LMVYKPAQNSMOQLADR 421
 QY 367 LVVRIRPSLVCCBGYIYAGG-DSVGGELNRTVERVYTEKDEMTVSPPLPCAMQMSAAV 425
 DB 422 LRCRGMADVAYLNGYIYILGGRDPTG-VKLKEVECYVQRNOMALVAPVPSFSLFDLM 480
 QY 426 VVHDCIYVMTLNLMYCYPPRSDSVEMAMRQTSRSPASAAFGDKIFYGGLHAIATNGI 485
 DB 481 VVRDIYALNSKRMFCYDPSHNMMLKCVSLKRN-DFOEACVFNEELIYICIDIPV----- 533
 QY 486 RLPSGTVDSSVTVEIYDVNKNEMMAANIPAKRYSDPCRAVVISNLSLQVNRETHANE 545
 DB 534 -----MKYVNPVRAEWRQNNIPLVSEITNN-VRIKHQKLLILTSRTPQWK 579
 QY 546 RAKYVTVQYDLELDKRWSLRQHISERV-LMDLGHDFCTGKLYPSGLSESPMKPPTLYFS 604
 DB 580 KKRVTYVEYDIREDDQWI---NIGTTLGLQPDNFPCLSNRVYPSCLF-----PQOSFL 630
 QY 605 TDGTE-----EFELDG 615
 DB 631 TEEBELPSESSTEWDLGG 648
 RESULT 9
 Q9H016 PRELIMINARY; PRT; 684 AA.
 AC Q9H016
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 77.2 kDa protein.
 GN DKFZP434E2318.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_Taxid=9606;
 RP [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=TESTIS;
 RX MEDLINE=21154917; PubMed=11330166;
 RA Wiemann S., Weill B., Wellenreuther R., Gassenhuber J., Glasel S.,

RA Ansgorge W., Boecker M., Bloecker H., Bauersache S., Blum H.,
 RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Struck N.,
 RA Mewes H.W., Oettermayr B., Obermaier B., Tampe J., Heubner D.,
 RA Wambut R., Korn B., Klein M., Poustka A.,
 RA "Towards a Catalog of Human Genes and Proteins: Sequencing and
 RT Analysis of 500 Novel Complete Protein Coding Human cDNAs."
 RL Genome Res. 11:422-435(2001).
 DR EMBL, AL136782; CAB66716.1; -
 DR InterPro: IPR000210; BTF POZ.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF00651; BTF; 1.
 DR Pfam: PF01344; Kelch; 3.
 DR SMART, SM00225; BTF; 1.
 DR PROSITE, PS50097; BTF; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 684 AA; 77153 MW; 9DD9273515A77555 CRC64;
 Query Match 19.2%; Score 630.5; DB 4; Length 684;
 Best Local Similarity 26.0%; Pred. No. 1.6e-44;
 Matches 176; Conservative 125; Mismatches 249; Indels 127; Gaps 21;
 QY 3 TQDERQINTEYAVSLLEQLKLYEQQLFTDIVLV-----EGTEPCHKMYLATCSS 54
 DB 35 TQPEELKDTAHSALLAQKSFYDARQLCDVTIEVTPGSGPTGRLPFCNNRVLAACP 94
 QY 55 YRAMPMSGLSESKQTHVLRNDATLQIITVATGNLANNDSVEQLYETACFLQVE 114
 DB 95 YKSMFTGMYESQOASVTMHVDVDAESFEVLVDYCTGVSLSEANVERLYAASDMLQLE 154
 QY 115 DYLRQCREYLKIKINAENCVRLLSPADLPSCSELKQSAKRWENHKTAAYH-----YHDAF 169
 DB 155 YRREACASFLARLDLTNCTAILKFPDAGHRLKRSQOASYAHQNFQSLSHMSIREETL 214
 QY 170 MQLSHDLIDLISDNLVNKEETVEAAMLMLENTESRQOYSSVLSQIRIDALSEVT 229
 DB 215 ADLTIAQLLAVALRLDSDVSEEGTCHVAVOMLEAPKRGPSAAEVFCVMMHPTBED 274
 QY 230 QRAMFGRL--PNDKSV---VVOG-----LYKSM----- 253
 DB 275 Q-DYLEGLTTRKYVKKYCLDVIIEGALQMRVYDLYSLVPVPNSSSSSSNSLVSAEN 333
 QY 254 -PKFPRKLGMTKEEMMIFLEASSENPCSLYSVCYSPQAEKYKLCSPPADLHKVTV 311
 DB 334 PQO---RLGMCCKEMVIFP-GHPRDP-----FLCDDYSGDLYKVPSPFLTCLAHRTVT 383
 QY 312 -----VTDNDIYIAGGOVPLKNTKTNHSTKSLQTAFTVNCFFWPAQONTWPKTPM 366
 DB 384 SSAVCSPDHDHDIYLA--QPRD-----LMVYKPAQNSMOQLADR 421
 QY 367 LVVRIRPSLVCCBGYIYAGG-DSVGGELNRTVERVYTEKDEMTVSPPLPCAMQMSAAV 425
 DB 422 LRCRGMADVAYLNGYIYILGGRDPTG-VKLKEVECYVQRNOMALVAPVPSFSLFDLM 480
 QY 426 VVHDCIYVMTLNLMYCYPPRSDSVEMAMRQTSRSPASAAFGDKIFYGGLHAIATNGI 485
 DB 481 VVQNYLVAVNSKRMFCYDPSHNMMLKCVSLKRN-DFOEACVFNEELIYICIDIPV----- 533
 QY 486 RLPSGTVDSSVTVEIYDVNKNEMMAANIPAKRYSDPCRAVVISNLSLQVNRETHANE 545
 DB 534 -----MKYVNPVRAEWRQNNIPLVSEITNN-VRIKHQKLLILTSRTPQWK 579
 QY 546 RAKYVTVQYDLELDKRWSLRQHISERV-LMDLGHDFCTGKLYPSGLSESPMKPPTLYFS 605
 DB 580 KKRVTYVEYDIREDDQWI---NIGTTLGLQPDNFPCLSNRVYPSCLF-----PQOSFL 631
 QY 606 -----DTEBEFLDG 615
 DB 632 EEDDARSSSTEWDLGG 648
 RESULT 10
 Q9NGX7 PRELIMINARY; PRT; 623 AA.
 ID Q9NGX7

AC G9NGX7;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Diablio.
 GN DBO OR CG6224.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA Stuart B.D., Wasserman S.A.;
 RT "Identification of Diablio, a new Drosophila melanogaster kelch family
 RT protein";
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF237711; AAF43447.1; -.
 DR FlyBase; FBgn0040230; dbo.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR01798; Kelch.
 DR Pfam; PF00651; BTB; 1.
 DR Pfam; PF01344; Kelch; 6.
 DR PRINTS; PRO0501; KELCHREPEAT.
 DR SMART; SM00225; BTB; 1.
 DR PROSITE; PSS0097; BTB; 1.
 SQ SEQUENCE 623 AA; 68951 MW; E65817EDEF4E7B6 CRC64;

Query Match 17.6%; Score 578.5; DB 5; Length 623;
 Best Local Similarity 27.4%; Pred. No. 3.5e-40;
 Matches 155; Conservative 103; Mismatches 215; Indels 93; Gaps 15;

18 LEOLKLFYEQQLFTDIVLIVGTEFPCHKVNLATCSSYFRAMFMSGLSESKQTHVHLRVV 77
 59 LTELNMRLRRRELCDVLANGRKIFAHRIYLSACSSYFCAMTGLSEBROQEVITRDI 118
 78 DAATLQIITVAYTGNLANMDSVTEOLYFACPLQVEDVLQRCREYLIKKNANCVRL 137
 119 DENAMELLIFCYTAHIVEESNVQTLPAACLLQVLEIODICEFELKQLDPTNCLGIR 178
 138 SPADLFSCEELKOSAKMVEHKFTAYVHODAFQGLSHDLIDILSDNINVEKEEVREA 197
 179 AFDTHSCRELHIADEFTQHNQOEWESEFLLPVGLVDIICDELNVRSDEQVFN 238
 198 AMLLEVTESRSQYSSVLSQIRIDALS-----EVTQRAFPG 236
 239 VMELKKNVNAERQHLAQVLRPLSPKFLVGTGSDLLVRSDEVCDDLVDKAKNVL 298
 237 LFPNDKSVVQGLYKSNPKFKKRLGNTKEEMMIFLEASSENPCSLYSVCSYSPQAEKY 296
 299 LTFQERF-LMQG--PRTRPR--KP-----TRGEVILFAVGCGSGDALASVERFDPQTNND-W 349
 297 KLSPPADLHKVGTVPDND-IYINGGVPLKNTKNSKTKLQTAFTVNCFFWEDA 355
 350 KMAVAFMSK-RRCGGAVNLMDLLYAAGG-----HGGSYLNSIER-----YDP 391
 356 QQNTW-FPKPTMFLVRIKPSLVCCEGYIYAGDSVGGELNRRTRVERDYDEKEMTWSP 414
 392 QTNOMSCDVAFPTTSCRSVGVAVLDGLYAVGGDQVCLNH--VERIDYDEKEMKSNVAP 449
 415 LFCAMQMSAAVVVHDCIYV-----TLNLMYCYPRSDSWEMAMRQTSRSPASAAAF 467
 450 MTRRLRGVAVAVLGGFLYAGSGDQCPLNTVERYPDRHKKVAVSPMSTRRLGCAVAF 509
 468 GDKIIFIG-----GLHAINSGRLPSGTVDS 495
 510 NNITVAVGGDDCMELSSAERYNPLTNMSPYAMTSRRSGVGLAVVNGQLYAVGGFDGS 569
 496 SV--TVEIYDVNKNEMKMANIPAKR 519
 570 AYLKTIETVDPETNOMRLCGCMYRR 595

RESULT 11
 ID Q9VU5 PRELIMINARY; PRT; 623 AA.
 AC Q9VU5;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG6224 protein.
 GN DBO OR CG6224.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxId=7227;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Suton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champs M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayeni A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Bencze P.V., Berman B.P., Bhattacharya D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mayhew A.D., Dew I., Dietz S.M.,
 RA Dodson G., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwa C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laso P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merilov G., Milhina N.V., Modarri C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Relvert K., Remington K., Saunders R.D.C., Scheller F., Shen H.,
 RA Shue B.C., Sideri-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Scapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.O.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195 (2000).
 DR EMBL; AE003529; AAF49578.1; -.
 DR FlyBase; FBgn0040230; dbo.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR001798; Kelch.
 DR Pfam; PF00651; BTB; 1.
 DR Pfam; PF01344; Kelch; 6.
 DR PRINTS; PRO0501; KELCHREPEAT.
 DR SMART; SM00225; BTB; 1.
 DR PROSITE; PSS0097; BTB; 1.
 SQ SEQUENCE 623 AA; 68923 MW; DCD75B9F66BDC6B9 CRC64;

Query Match 17.6%; Score 578.5; DB 5; Length 623;
 Best Local Similarity 27.4%; Pred. No. 3.5e-40;
 Matches 155; Conservative 103; Mismatches 215; Indels 93; Gaps 15;

18 LEOLKLFYEQQLFTDIVLIVGTEFPCHKVNLATCSSYFRAMFMSGLSESKQTHVHLRVV 77

Db 59 LTELANKRRLRELCDDVAVLVGKRIKPAHVIISACSSFCAMFTGELBSRQTEVTRDI 118
Qy 78 DAATQIIITTAATGNLANNDSTVEQLYETACFLQVEDVYQRCREYLKIKINAEVCRL 137
Db 119 DEANAMELLIDFCYTAHIIVEESNVQTLPLPAACLLQVEI IDICCEFLKQDLPTNLCGR 178
Qy 138 SPFADLFSCGELKQSAKRWHEKHTAAYHODAFMQLSHDLIDLSDNLNVEKEEYVRA 197
Db 179 APADTHSCRELLRIADKFTQHNQFWESEEFLLLEVGLVDLICDELVNSEEQVFA 238
Qy 198 AMULEYNTESRQYLSVLSQIRIDALS-----EYVQRAMFG 236
Db 239 VMSMLKYNABERKQHLAQVLOHRLPLSPKFLVGTGSDLVRSDEACRDLVDEAKNYL 298
Qy 237 LPENDSVVVOGLYKMPKFKRQLGTMTEEMMIFLEASSENPCSLYSSVCSPQAEKY 296
Db 299 LIPQER- LMOG- PRTRPR- KP-----TRGEVLFAVGCGSDALASVBRFPQND- W 349
Qy 297 KLCSPPADLHKVGTVTTPDND- IYAGGVPLKNTKTNSTKSKLQTAFTTNCIFYWPA 355
Db 350 KVAAPMSK- RRCGVAVLNDLLYAVG-----HDGQSYLNSIER-----YDP 391
Qy 356 QONTW- FRKTPMLFVRIKPSVCEGYIYAGSDSVGELNRTVERYPTEKDEMTWSP 414
Db 392 QTNQMSCDVAFPTTSCRTSVGAVLDGFLYAVGSDGVQCLNH- VERYPKEMKSKVAP 449
Qy 415 LPCAMQSAVAVVHDCIYVW-----TLNMYCYPREDSVWEMAMROTSSRFSASAAP 467
Db 450 MTRTRRGVAVVAVLGGFLYVAGSDGQCLNTVERYPKRNKVAWSPMSTRHGLCANF 509
Qy 468 GDKIPIFG-----GLHIATNSGIRLPSGTVDG 495
Db 510 NNYIYAVGGRDQCMELSSAERYNPLTNTWSPIVAMTSRRSGVLAVVNGQLYAVGFGDS 569
Qy 496 SV--TVEIYDVNKEWMAANIPAKR 519
Db 570 AYLKTIETVDPETNOMRLCGCMYRR 595

RESULT 12

08VCK5
ID 08VCK5 PRELIMINARY, PRT, 591 AA.
AC 08VCK5;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Hypothetical 65.9 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019571; AAH19571.1; -
DR InterPro; IPR000210; BTR POZ.
DR InterPro; IPR001798; Kelch.
DR InterPro; IPR001005; MYD_DNA_binding.
DR Pfam; PF00651; BTR 1.
DR Pfam; PF01344; Kelch; 6.
DR PRINTS; PRO0501; KELCHREPEAT.
DR SMART; SM00225; BTR 1.
DR PROSITE; PS00097; BTR 1.
DR PROSITE; PS00334; MYB_2; UNKNOWN_1.
KW Hypothetical protein.
SQ SEQUENCE 591 AA; 65939 MW; D4A737FAD7F1A2EB CRC64;

Query Match 17.2%; Score 565.5; DB 11; Length 591;
Best Local Similarity 26.3%; Pred. No. 4, 1e-39;
Matches 151; Conservative 109; Mismatches 222; Indels 93; Gaps 14;

Qy 9 INTERAVSLLEQLKIFYEQQLFTDVLIVETGEPFCHKMTATGCSYFRAMMSGLSBSK 68
Db 28 IDKQPROTELEINILRKRLRELCDDVAVLVGAKKIPAHVIISACSSFCYFRAMFTGELBSR 87
Qy 69 QTHVLRNVDAATLQIIITTAATGNLANNDSTVEQLYETACFLQVEDVYQRCREYLKIKI 128
Db 88 QTEVAVTRIDERAMELLIDFATTSQIYBEGNVQTLPAACLLQVEI IDICCEFLKQDLPTNLCGR 147
Qy 129 NAEVCRLSPFADLFSCGELKQSAKRWHEKHTAAYHODAFMQLSHDLIDLSDNLNVEKEEYVRA 188
Db 148 DSNLCIGRAFDTHSCRELLRIADKFTQHNQFWESEEFLLLEVGLVDLICDELVNSEEQVFA 207
Qy 189 EKEETVREANMULEYNTESRQYLSVLSQIRIDALS-----E 227
Db 208 RSEEQVFAVMAWVYKSIQERRPOLPOVLOHRLPLSPKFLVGTGSDLVRSDEACRDLVDEAKNYL 267
Qy 228 VQRAMFGQLPENDSVVVOGLYKMPKFKRQLGTMTEEMMIFLEASSENPCSLYSSVCSPQAEKY 284
Db 268 LVDEAKNYLLIPQER- LMOG- PR- TRPKRCIRGEVLFAVGW---CSGDAISS 316
Qy 285 SVCYSPQAEKYKLCSPPADLHKVGTVTTPDNDIYAG-----QVPLKNTKTNH-- 334
Db 317 VERYPQNTNEMWMAWMSKRGCGVSVL- DLLAVGCHDSSLYLNSVERYPDPTNOMS 375
Qy 335 ---SKTSKLQTA-----FTTNCIFYWPAQONTWEPKTPMLFVRIRK 373
Db 376 SDVAFPTTSCRTSVGAVLGGFLYAVGSDGVQCLNIVERYPKEMKSTRRLCGV 435
Qy 374 SLVCEGYIYAGSDSVGELNRTVERYPTEKDEMTWSPPLPCAMQSAVAVVHDCIY 433
Db 436 AVAVLGGFLYAVGSDGTSPLN- TVERNPEENMHTIAPGTRRKHLGCAVYODMTYA 493
Qy 434 M-----TLNMYCYPREDSVWEMAMROTSSRFSASAAPFQDKIFYIGLHIATNSGIR 486
Db 494 VGGRDITELSSAERYNPLTNTWSPIVAMTSRRSGVLAVVNGQLYAVGFG----- 544
Qy 487 LPSGTVDGSSV--TVEIYDVNKEWMAANIPAKR 519
Db 545 -----DGTTLKTIETVDPETNOMRLCGCMYRR 573

RESULT 13

08TBH5
ID 08TBH5 PRELIMINARY, PRT, 593 AA.
AC 08TBH5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Kelch-like 2, Mayven (Drosophila).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA Strausberg R.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC022503; AAH22503.1; -
SQ SEQUENCE 593 AA; 66075 MW; 3000E53B69E57F0D CRC64;

Query Match 16.3%; Score 537; DB 4; Length 593;
Best Local Similarity 25.1%; Pred. No. 1, 1e-36;
Matches 136; Conservative 101; Mismatches 204; Indels 100; Gaps 13;

Qy 5 DERQINTB-----YAVSLLEQLKIFYEQQLFTDVLIVETGEPFCHKMTAT 51
Db 20 DSKDNTNTEGHCFTVTPMWMKKAFFKVMNELR---SQNLICDVTIVAEDMEISAHVVLAA 76
Qy 52 CSYFRAMMSGLSBSKQTHVLRNVDAATLQIIITTAATGNLANNDSTVEQLYETACFL 111
Db 77 CSPYHAFMTGMSBSRRAKRVRIKEVDGWTLMILIDVYVTAIOVTEENVOVLLPAGLL 136

QY 112 QVEDVLORCREYIKKINAEVCVRLSPADLFSCEELKQSAKRMVEHKFTAVYHDAFMO 171
 Db 137 QLODVKTCCCEFLSPQHPVNCGLGIRAFADMAHACTDLNKAANTYAEQHFADVLSEEFNL 196
 QY 172 LSHDLLIDILSSNINLVEKETREAMLMLEVENTESRSQYLSVLSQIRIDALSE--VT 229
 Db 197 LGIEOVCSLISDGLTISSEKVEAVIAVNNHDKDROEFMARLMEHVLPLPREYLV 256
 QY 230 QRAMFOGLPND--KSVVQGL-YKSMPEFKPRLQMTKEBMIFLEASSENPCSLYS 285
 Db 257 QRVEEALVKNSSACDYILEAMKHYLP-----TEQRILM----- 292
 QY 286 VCSFPAQEKYKLCSPPADLHKVGTVPNDIYIAGGVPLKNTTNSKTSKLOTAER 345
 Db 293 -----KSVTRLRTIPMLPKLMVVV-----GGQAP-----KAIR 321
 QY 346 TVNCFWPDQONTWPKFPMFLVRIKPSLVCEGYIYAGDSVGGELNRTVERDYDE 405
 Db 322 SYEC---YDFKERMHOVAELPSRCRCAQVYMAGLVFAVG--FNGSLKVRTVDSDYDV 376
 QY 406 KDEMTWVSLPCAMQMSAAVVVHDCIYVM-----TLNLMCYFPRSDSWEMAROTS 458
 Db 377 KOQMTSVANMRDRSTLGAANGLYAVGFGDSTGLSVEAYNIKSNMFHVAEMNTR 436
 QY 459 RSPASAAAFGDKIFYIGGLHIATNSGIRLPSGTVDSSVTEIYDVNKEKMAANI 518
 Db 437 RSSVGVGVGGLLYAVGVDAVSRCGLS-----TVECYNATYEMTYIAEMSTR 485
 QY 519 R 519
 Db 486 R 486

RESULT 14

Q9BQF8 PRELIMINARY; PRT; 586 AA.
 AC Q9BQF8; 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Hypothetical 66.0 kDa protein.
 GN DKFZ564C16.6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA MEDLINE=21154917; PubMed=11230166;
 RA Wiemann S., Weill B., Wellenreuther R., Gaessenhuber J., Glassl S.,
 RA Ansoyge W., Boescher M., Bloeker H., Bauersachs S., Blum H.,
 RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
 RA Mewes H.W., Ottenwelder B., Obermaier B., Tampe J., Heubner D.,
 RA Wandt R., Korn B., Klein M., Poulet A.;
 RA "Towards a Catalog of Human Genes and Proteins: Sequencing and
 RA Analysis of 500 Novel Complete Protein Coding Human cDNAs.";
 RL Genome Res. 11:422-435 (2001).
 DR EMBL: AL136597; CAB6532.1; --
 DR HSSP: Q05516; ICS3.
 DR InterPro: IPR000210; BTB_POZ.
 DR InterPro: IPR001798; Kelch.
 DR Pfam: PF00651; BTB; 1.
 DR Pfam: PF01344; Kelch; 5.
 DR SMART: SM00325; BTB; 1.
 DR PROSITE: PS50097; BTB; 1.
 DR Hypothetical protein.
 SQ SEQUENCE 586 AA; 65991 MW; FA4ED8F1567AC7A8 CRC64;

Query Match 16.0%; Score 526.5; DB 4; Length 586;
 Best Local Similarity 25.7%; Pred. No. 8; 1e-36;
 Matches 156; Conservative 100; Mismatches 229; Indels 123; Gaps 20;

QY 11 TEVAVSLLEQLT-----FYEQGLFTDILVIEGTEPPCKRMVLATCSSYFRAME 60
 Db 14 TEKKLAAREBAEKLFGFMVNMNRKQKTLCVIILMVERKIPARHVVLAAASHPEFLMF 73
 QY 61 MGGLSEKQTHHLNNDATLQIIITVYATGNLMDNSTVQYETACFLQVEVQLQC 120
 Db 74 TTNLMESKSFVELDKDAEDDILEQVEFAYTARISVNSNNVSLDDAQAQYIEPVKMC 133
 QY 121 REYLTKKINAENCVRLLSPADLFSCEELKQSAKRMVEHKFTAVYHDAFMOQLSHDLLDI 180
 Db 134 VFLEKQVDASNCGLGVLAECDCPELKATDDPIHQHFFTEVYTDDELQIDVRYTHL 193
 QY 181 LSSDNLNVEKETREAMLMLEVENTESRSQYLSVLSQIRIDALSE----- 227
 Db 194 LMQDTLYVAEDQVDAVRLKLYDEPNRQPPMVILAKVRPLSLKSTQVAEPPI 253
 QY 228 -----VTQRAPFOGLPNDKSVVQGLYKSMPEFKPRLQMTKEBMIFLEASSEN 278
 Db 254 QDNPECLKXVIGMRVHLSPEDRELVDT-----RPRKRGDYRIALF-GSQOPQ 304
 QY 279 PCSLYSVCSYSPQAEKVKYLCSPPADLHKVGTVPNDIYIAGQ--VPLKNTKTNHSHK 336
 Db 305 SCRYENPMDYS-----WTDIRCPFEKRRDAACVMDNVYILGSSQLFPIR----- 351
 QY 337 TSKLOTAFTVNCFYWFDQONTWF---PKTPMLFVRIKPSLVCC--EGYIYALIGDSV 390
 Db 352 -----MDC---YNNVVDKSWYSKLGPPTR-----RDLAACAAGKITYSGSSEV 392
 QY 391 GSELNRRYVERDYDEKDER-TWVSLP--CAMQMSAAVVVHDCIYV-----MTL 436
 Db 393 GNSA-LYLFECYDITESTWHTKPSMLTORCS--HGWENGLIYVCGSLANNVSGRYL 448
 QY 437 NLMCYFPRSDSWEMAMAROTSRSFASAAAFGDKIFYIGGLHIATNSGIRLPSGTVDGSS 496
 Db 449 NSCEYVDATETWTBLCPMIEARKHGLVYVVDKIFAVCG-----QNGL-----GGL 495
 QY 497 VTVEIYDVNKEKMAANIIPAKRYSDPCVRAVVISNLCVEN--RETHLNERAKYVYQ 553
 Db 496 DNVVEYDIDKLNEMKQVSPMPKGVTKCAAVGSIYVLAGFGVGLGHILE----- 547
 QY 554 YDLLELDRW 561
 Db 548 YNTEIDKW 555

RESULT 15

Q9VENS PRELIMINARY; PRT; 744 AA.
 AC Q9VENS; 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CG3962 protein.
 GN CG3962.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouch J., Brokstein P., Brothier P.,

RA Burtis K.C., Bugam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.E., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dong L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harits N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hoslin D., Houshun K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasok P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mlehnina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Maasman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL; AF003716; AAF55386.1; -;
DR Flybase; FBgn038475; CG3962.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF01344; Kelch; 6.
DR PRINTS; PR00501; KELCHREPEAT.
DR SMART; SM00225; BTB; 1.
DR PROSITE; PS50097; BTB; 1.
SQ SEQUENCE 744 AA; 83604 MW; 6F7A4EB69DF482E9 CRC64;

Query Match 16.0%; Score 526.5; DB 5; Length 744;
Best Local Similarity 28.0%; Pred. No. 1.2e-35;
Matches 157; Conservative 77; Mismatches 207; Indels 119; Gaps 18;

QY 11 TEVAVSLLEQLKLYEQQLFTDVLIVEGTEPFCNKLATCSSYFRAMFMSGLSESQOT 70
DB 38 SNVAKEALQKMTMMRSHGLTDLVLEKKELFPAHKVVLSSAASPYKAMFTGSLKESMS 97
QY 71 HVHLRNVDAATLQIITTAAYTGNLANNDSTVEQLYETACFLQYEDVLRQREYLKIKINA 130
DB 98 RVQLQGVCPATMSRLIYFMVYTGQIRVTEVTCQLLPAAATMFQVPNVIDACCAFLERQIDP 157
QY 131 ENCVRLLSFADLPSCSELKQSAKRWVHEKFTAVYHODAFQLSHDLIDILSDNLNVEK 190
DB 158 TNAIGIAHFAEOHCVELQKRAVFLERNFTQYQOESEFLQLSAYQLIALIRDELDNVOE 217
QY 191 EETVREAAMLMEYNTESRSQYLSVLSQIRIDALSEVTGRAMFQGLPPNDKSVVVOGLY 250
DB 218 EREVVAVAVLKWKYDSDNHCMEHILGAVRC-----QPLTPN----- 255
QY 251 KSMPEKFKRLGKTKEMMI FTEASENPCLSYSV--CYSPOAEKRVYKLCSPADLHKV 308
DB 256 -----FLKEQM-----KNCADVLRKVPACREYLA-KIFK-----DLTLHKC 289
QY 309 GTYV--TPND--IYAG-----QVPLKTKTKNHSKTSKLTQAF-- 344
DB 290 PGVKERTPTMTMTI FVAGGFRHSLDILBAVNDVDMTWTTLANLRI--PRSLGAAPLK 346
QY 345 -----RTVNCFY--WPD--AQONTWPKPTMLFVRIKPSLVCCGYYIYAGD 388
DB 347 GKPYAVAGGRNNNGSSYDSQWVDRYSAVIETWPCAPMSVPRHRVGVAVMDELMTAVGG- 405
QY 389 SVGGELNRRITVERIDTEKDEWTVSP LPCAMQMSAAVVVHDCIYVM-----TLNLMYC 441

DB 406 SAGMEYH-NTVEYIDPDDLRWTLVOPMFAKRLGCVVVVNNLLVAIGFPDGNERLASVBC 464
QY 442 YFPRSDSVWEMAMRQTSRSPASAAAFGDKIFYIGLHATNSGIRLPGTVDGSS--VTV 499
DB 465 YHPENNEMSPLEPLDTGSGAGVAALNGYIYVVGFS-----DGTQLATV 509
QY 500 EYDVNKNEMKQANIPAKR 519
DB 510 ERYDTENDTWMVAPIQIAR 529

Search completed: July 14, 2003, 18:20:53
Job time : 33.9026 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:10:22 ; Search time 11.4598 Seconds
(without alignments)
4345.419 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789
Sequence: 1 MGSPAPGALGYREFTNRH.....RQKGAATNTKVHILGAP 518

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2556	91.6	480	2	JC7812 BCL6 homologous zi
2	1128.5	40.5	706	2	152586 B-cell CLL/lymphom
3	1122.5	40.2	706	2	A48752 B-cell CLL/lymphom
4	559	20.0	794	2	S59069 Z13 protein - mous
5	482	17.3	610	2	JC7315 myoneurin - human
6	481	17.2	688	2	A56360 zinc finger protei
7	434.5	15.6	614	2	JH0500 zinc finger protei
8	432.5	15.5	555	2	153869 zinc finger protei
9	424	15.2	595	2	G02075 transcription repr
10	421.5	15.1	673	2	S36336 probable transcrip
11	419	15.0	676	2	150643 gammaFp-C - chick
12	417.5	15.0	701	2	T14757 hypothetical prote
13	415	14.9	803	2	S26823 zinc finger protei
14	413.5	14.8	828	2	A32891 zinc finger protei
15	412.5	14.8	693	2	137570 zinc finger protei
16	408	14.6	728	2	A48830 probable transcrip
17	407	14.6	209	2	S47068 finger protein HZF
18	404	14.5	668	2	T08725 probable finger pr
19	401	14.4	1191	2	S35305 zinc finger protei
20	399	14.3	469	2	136600 zinc finger protei
21	399	14.3	485	2	A40751 finger protein MZF
22	397	14.2	543	2	B34612 zinc finger protei
23	396	14.2	449	2	S41647 zinc finger 5 prot
24	394.5	14.1	594	2	T12488 hypothetical prote
25	394	14.1	196	2	S06561 finger protein (cl
26	393.5	14.1	466	1	TMFP transcription fact
27	393	14.1	475	2	S03679 finger protein (cl
28	392	14.1	292	2	S43826 finger protein OZF
29	390	14.0	393	2	JN0533 finger protein pML

30	389.5	14.0	195	2	S00754 zinc finger protei
31	389.5	14.0	733	2	JC7679 dendritic cell-der
32	387.5	13.9	553	2	S22954 finger protein zfp
33	387.5	13.9	675	2	S51037 zinc finger protei
34	387	13.9	325	2	138616 zinc finger protei
35	387	13.9	448	2	S71427 transcription fact
36	386.5	13.9	580	2	A37107 spermatogenesis pr
37	386	13.8	654	2	A57785 finger protein ZNF
38	385	13.8	572	2	139311 Kruppel-type zinc
39	384.5	13.8	347	2	S00549 developmental cont
40	383.5	13.8	169	2	A39240 finger protein mfg
41	382.5	13.7	540	2	B57785 zinc finger protei
42	381	13.7	201	2	157505 zinc finger protei
43	380.5	13.6	247	2	S47070 finger protein HZF
44	380.5	13.6	261	2	S70006 finger protein zfo
45	380.5	13.6	710	2	148668 zinc finger protei

ALIGNMENTS

RESULT 1

JC7812 BCL6 homologous zinc finger protein BAZF - human
C:Species: Homo sapiens (man)
C>Date: 03-Jun-2002 #sequence_revision 03-Jun-2002 #text_change 03-Jun-2002
C:Accession: JC7812
R:Sakashita, C.; Fukuda, T.; Okabe, S.; Kobayashi, H.; Hirosewa, S.; Tokuhisa, T.; Miyas
Biochem. Biophys. Res. Commun. 291, 567-573, 2002
A>Title: Cloning and characterization of the human BAZF gene, a homologue of the BCL6 on
A:Reference number: JC7812; PMID:11855826; MIMD:21845446
A:Accession: JC7812
A:Molecule type: DNA
A:Residues: 1-480 <SAK>
A:Cross-references: DDBJ:AB076580; DDBJ:AB076581
C:Comment: This protein, a zinc finger protein containing a conserved amino terminal BTB
rly in the induction of megakaryocytic differentiation to produce mature platelets. This
cell fate.
C:Genetics:
A:Gene: bazf
A:Map position: 17p13.1
A:Introns: 60/2; 134/2; 256/2; 297/3; 353/1; 396/3; 442/3
C:Keywords: differentiation

Query Match	91.6%	Score 2556	DB 2	Length 480
Beef Local Similarity	92.7%	Pred. No. 3.5e-137		
Matches	480	Conservative	0	Mismatches 0; Indels 38; Gaps 1
QY	1	MGSPAPGALGYREFTNRHSDVLTGNTLRLGILTDVTLVGGCPRLRAKAVLIACS	60	
DB	1	MGSPAPGALGYREFTNRHSDVLTGNTLRLGILTDVTLVGGCPRLRAKAVLIACS	60	
QY	61	GFYISIFRGAGVGVVLSLPGGPARGPAPLDMYTSRLSLSPATAPAVLAATYLOM	120	
DB	61	GFYISIFRGAGVGVVLSLPGGPARGPAPLDMYTSRLSLSPATAPAVLAATYLOM	120	
QY	121	EHVVOACHRFQASVEPGLISLRPLEAPPTPTAPPGSPRRSGHDPPTESCSOG	180	
DB	121	EHVVOACHRFQASVEPGLISLRPLEAPPTPTAPPGSPRRSGHDPPTESCSOG	180	
QY	181	PPSPAPDPKACNMKKYTYILNSQASQASGLVGRSSGQPCQARLSPGDEASSSSSSS	240	
DB	181	PPSPAPDPKACNMKKYTYILNSQASQASGLVGRSSGQPCQARLSPGDEASSSSSSS	240	
QY	241	SSSSSEEGPIPGQSRSLSPATATVQFKGAPASTPYLLTQAODTSGSPSEARLPQSE	300	
DB	241	SSSSSEEGPIPGQSRSLSPATATVQFKGAPASTPYLLTQAODTSGSPSEARLPQSE	300	
QY	301	PPSCNCEAVAGCSGLSLVPGDEDKPKYKQOLCSSFRYGNLASHRTYHTGKPYHCS	360	
DB	301	PPSCNCEAVAGCSGLSLVPGDEDKPKYKQOLCSSFRYGNLASHRTYHTGKPYHCS	360	
QY	361	ICGARFNPANLKTSHRIHSGEKPKYKCTCGSRFQVARSQPPSGFQKPARGCVQKGGF	420	

```

Db 361 ICARFRRPANTKTHSHIGGEKPKETCGSRFV----- 395
Qy 421 CSSGRODLKSPQSOVLAHRAVLIHTGEKPYPCPTGTRPRHLOTLKSHVRIHTGKPYH 480
Db 396 -----QVALHRAVLIHTGEKPYPCPTGTRPRHLOTLKSHVRIHTGKPYH 442
Qy 481 CDPCGLHFRHKSQRLRLHROKGAATNTKVYHILGSP 518
Db 443 CDPCGLHFRHKSQRLRLHROKGAATNTKVYHILGSP 480

RESULT 2
B-cell CLL/Lymphoma 5 (BCL5) protein - human
N:Alternate names: finger protein LAZ-3; zinc finger transcription factor BCL-5
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence, revision 01-Nov-1996 #text_change 31-Dec-2000
C:Accession: I52586; S40520; S32767
R:Miki, T.; Kawamata, N.; Hirohata, S.; Aoki, N.
Blood 83, 26-32, 1994
A>Title: Gene involved in the 3q27 translocation associated with B-cell lymphoma, BCL5.
A:Reference number: I52586; MUID:94100541; PMID:8274740
A:Accession: I52586
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-706 <MIK>
A:Cross-references: GB:867779; NID:9459372; PIDN:AAB29544.1; PID:9459373
R:Kerchaert, J.P.; Dewelindt, C.; Tilly, H.; Quief, S.; Lecocq, G.; Bastard, C.
Nature Genet. 5, 66-70, 1993
A>Title: LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3
A:Reference number: S40520; MUID:94035122; PMID:8220427
A:Accession: S40520
A:Molecule type: mRNA
A:Residues: 1-706 <KER>
A:Cross-references: EMBL:Z21943; NID:9297025; PIDN:CAA79937.1; PID:9297026
C:Genetics:
A:Gene: GDB:BCL5; LAZ-3; ZNF51
A:Cross-references: GDB:125178; OMIM:151441; GDB:138176; OMIM:109565
A:Map position: 17q22-17q23; 3q27-3q27
A:Introns: 60/3
C:Superfamily: POZ domain homology
C:Keywords: DNA binding; transcription factor; zinc finger
F:18-121/Domain: POZ domain homology <POZ>
F:520-541/Region: zinc finger CCHH motif
F:548-568/Region: zinc finger CCHH motif
F:576-596/Region: zinc finger CCHH motif
F:604-624/Region: zinc finger CCHH motif
F:632-652/Region: zinc finger CCHH motif
F:660-681/Region: zinc finger CCHH motif

Query Match 40.5%; Score 1128.5; DB 2; Length 706;
Best Local Similarity 35.0%; Pred. No. 156-56;
Matches 270; Conservative 48; Mismatches 117; Indels 337; Gaps 16;

Qy 1 MGSPAPBEGALGVREFTRHSDVLGNLRLRGILITDVTLLVGGQPLRAKRAVLIACS 60
Db 1 MASPA--DSCI-----QFTRHNSDVLNLNRLRSRDILTDVIVVSRBQFRAKTVLMACS 54
Qy 61 GFYFSTFRGAGVGVVLSLPGSPBAGAPPLDDPWYTSLSRLSPATAPVLAATYLLQ 120
Db 55 GLFYSTFTQDLKCNLSVINDPINEGFCILLDFWYTSLNLRGNIMAVMATAYLLQ 114
Qy 121 EHVYQACHRFIOAS----- 134
Db 115 EHVYDTCRKRTKASEAMVSAIKPRBEFLNSRLMLPQDITMAYGRGVENNLLPLASAPG 174
Qy 135 -----YEPLAGIS----- 141
Db 175 CESRAPAPSLVSGSLSTPPASYSWYSHLPVSGLLFSDBEFVDVMPVNPPEKRALPCDS 234
Qy 142 -----LRPL----- 145

```

```

Db 235 ARPVGESRPTLEWSPNVCHSNISPKETIBEARSDMHWAVAGLKPAPASARNAPYF 294
Qy 146 -----EAEPPT-----PTAP-----PGSPRSEGGPDPTESRS----- 176
Db 295 PCDKASKEERFSSDEELHPEFPNAPLNKGLVSPGSPQSDQPNPTSSCSKAC 354
Qy 177 ---CSGPPSPASPPEKACNMKKYKIVLNSQASQAGSLVGRSSGQCPQARLPBGDEA 233
Db 355 ILQASGSPPAKSPTPBKAACNMKKYKIVLN----- 384
Qy 234 SSSSSSSSSSEEGPIPPQSRILSPATATVPKCGAP-----ASTPYLLTQAQDT- 285
Db 385 ----SLNQNAKPEGEQALGRLSPRAYTAPACQPMEPENLDLQSPFKLSAGEDST 439
Qy 286 -----SGSP-----SERAPY----- 296
Db 440 IQASRLNNIVRSMTHSGSRSSSESHSLPYHMPKCTSGSGSPQHAEMCLHTAGPTPE 499
Qy 297 ----PGSEF-----FSCQNCBAVAGCSSGLD-SLVPGDEKPYKQCLCRSFRYKG 342
Db 500 EMGETQSEYSDSCENGAFPCNECDRCFSEASLKRHTLQTHSDRPYKCDRCQASFRYKG 559
Qy 343 NLASHRTVYTGKPYHCSICGARFNRPNALKTHSHIGGEKPYKCTGSRFVQVRSQRP 402
Db 560 NLASHRTVYTGKPYRCNLCGAQFNRPAVLKTHRIHSGEKPYKCTCGARFV----- 612
Qy 403 SGFGKPARGAGVGQKGFCSQRODLKSPQSOVLAHRAVLIHTGEKPYPCPTGTRPRH 462
Db 613 -----QVALHRAVLIHTGEKPYPCPTGTRPRH 641
Qy 463 LQTLKSHVRIHTGEKPYHCDPCGLHFRHKSQRLRLHROKGAATNTKVYH1 514
Db 642 LQTLKSHVRIHTGEKPYHCEKCNLHFRHKSQRLRLHROKGAATNTKVQYRV 693

RESULT 3
A48752
B-cell CLL/Lymphoma 6 (BCL6) protein - human
N:Alternate names: zinc finger transcription factor BCL-6
C:Species: Homo sapiens (man)
C:Date: 07-Oct-1994 #sequence, revision 07-Oct-1994 #text_change 01-Dec-2000
C:Accession: A48752
R:Ye, B.H.; Lieta, F.; Cogo, F.L.; Knowles, D.M.; Offit, K.; Chaganti, R.S.K.; Dalla-Fav
Science 262, 747-750, 1993
A>Title: Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphc
A:Reference number: A48752; MUID:94053709; PMID:8235596
A:Accession: A48752
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-706 <YEA>
A:Cross-references: GB:U00115; NID:9392426; PIDN:AAC50054.1; PID:9392427
C:Genetics:
A:Gene: GDB:BCL6; BCL5; LAZ3; ZNF51
A:Cross-references: GDB:138176; OMIM:109565
A:Map position: 3q27-3q27
C:Superfamily: B-cell CLL/Lymphoma 5 protein; POZ domain homology
C:Keywords: transcription factor; zinc finger
F:18-121/Domain: POZ domain homology <POZ>
F:520-541/Region: zinc finger CCHH motif
F:548-568/Region: zinc finger CCHH motif
F:576-596/Region: zinc finger CCHH motif
F:604-624/Region: zinc finger CCHH motif
F:632-652/Region: zinc finger CCHH motif
F:660-681/Region: zinc finger CCHH motif

Query Match 40.2%; Score 1122.5; DB 2; Length 706;
Best Local Similarity 36.3%; Pred. No. 336-56;
Matches 268; Conservative 58; Mismatches 142; Indels 271; Gaps 17;

Qy 1 MGSPAPBEGALGVREFTRHSDVLGNLRLRGILITDVTLLVGGQPLRAKRAVLIACS 60
Db 1 MASPA--DSCI-----QFTRHNSDVLNLNRLRSRDILTDVIVVSRBQFRAKTVLMACS 54

```

QY 61 GFFYSIFRGRAGVUDVLSLPGGEPKAGAPLIDMYTSTRRLSPATAPAVLAATYLOM 120
| | | | |
DB 55 GLPYSIFDTQKCNLSVINLDEINDEGCILIDFMYSTRMLNREGINMAVATATYLOM 114
| | | | |
QY 121 EHVYQACHRFQAS----- 134
| | | | |
DB 115 EHVVDTCRKFICASEAMVSAIKPREEPLNSRMLPQDIMAYRGVEVNNILPLRSADG 174
| | | | |
QY 135 -----YEPGLGIS----- 141
| | | | |
DB 175 CESRAFAPSLYGSLSTPPASYSMTSLPVSLLFSDSEFRDVRMPVANDPPEKRALPCDS 234
| | | | |
QY 142 -----LRPL----- 145
| | | | |
DB 235 ARPVEYSPTLEVPNVCHSNISPKETIPEARSDMHVSVAEGLKPAASARAPYF 294
| | | | |
QY 146 -----EAPPT-----PPTAP-----PPGSPRSEGHDPPTESRS----- 176
| | | | |
DB 295 PCDKASKEERPSSEDEIALHFEPPNAPLNKGLVSPQSPQSDCQPNSTPCSSKXAC 354
| | | | |
QY 177 ---CSGPPSPASPDPKACNMKKYKTYLVNS-----QASQAGSLVGERSSGQPCP 223
| | | | |
DB 355 IIOAGSPSPAKSPTPDKACNMKKYKTYLVNSLQNAKPGPEQAEIGRLSPRAYTPPAC 414
| | | | |
QY 224 QARLPSEDEASSSSSSSSSESPIPGPQSF-----SPTAATVQF----- 266
| | | | |
DB 415 QPPM--EPENLDLQSPFKLASGEDSTIP--QASRLNNIYNRSMTGSPRSSSHSPLYMAP 472
| | | | |
QY 267 ---KCG--APASTPYLLTSQ---AODTSGSPSEARPLPGESEFSSCONCEVAGCSSG 316
| | | | |
DB 473 PCTSGSGSPGHAEMCLHTAGPTFAEMGETOSEYSDSSCENGAFPCNECDRFESEAS 532
| | | | |
QY 317 LD-SLVPGDEBQPYKQLCRSSSPRYKGNLASHRTVTGGEKYHCISICGARFNRPANLKT 375
| | | | |
DB 533 LGRHILQTHSDRKYKCDRCQASFRYKGNLASHKTVTGGEKYRCNLCGAQFNRPANLKT 592
| | | | |
QY 376 SRHSGEKPCKETCGSRFVQVRSPPSGFQGRAGVQKQGFCSQRDLKSPSOV 435
| | | | |
DB 593 TRHSGEKPCKETCGARFV----- 614
| | | | |
QY 436 AHLRAVLLHTGEKPRPCCTGFRRLQTLKSHVRIHTGEKPYHCDPCGLHFRHKSQR 495
| | | | |
DB 615 AHLRAVLLHTGEKPRPCCTGFRRLQTLKSHVRIHTGEKPYHCEKCNLHFRHKSQR 674
| | | | |
QY 496 LHLRQKGAATNTKVYHI 514
| | | | |
DB 675 LHLRQKGAATNTKVYRV 693
| | | | |
RESULT 4
559069
C1:Species: Mus musculus (house mouse)
C1:Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 28-May-1999
C1:Accession: S59069
R1:Schulz, T.C.; Hopwood, B.; Rathjen, P.D.; Wells, J.R.B.
Biochem. J. 311, 219-224, 1995
A1:Title: An unusual arrangement of 13 zinc fingers in the vertebrate gene Z13.
A1:Reference number: S59069; PMID:96003919; PMID:7575457
A1:Accession: S59069
A1:Status: preliminary
A1:Molecule type: mRNA
A1:Residues: 1-794 <SCH>
A1:Cross-references: GB:U14556; NID:9608136; PIDN:AAA5493.1; PID:9608137
A1:Superfamily: POZ domain homology
F1:0-108/Domain: POZ domain homology <POZ>
Query Match 20.0%; Score 559; DB 2; Length 794;
Best Local Similarity 26.6%; Pred. No. 2,2e-24;
Matches 164; Conservative 73; Mismatches 198; Indels 182; Gaps 17;
QY 16 EFRHSSDVLGNINELRLRGILTDVTLVGGQPLRAHKAVALACGFPYSIFRGRAG-VG 74
| | | | |

DB 2 DFRQHSQRLVEQLNQORQLGICDCTFVVDGVDFRAHKAVALAACEYRMLFVDQKDVH 61
| | | | |
QY 75 VDVLSPGGEPRARGAPLIDFMYTSTRRLSPATAPAVLAATYLOMEHVYQACHRFQAS 134
| | | | |
DB 62 LDI-----SNAAGIGQVLEFFYTKLSLSPENVDVLAVASFLQMODIVTACH-LKSL 114
| | | | |
QY 135 YEPGLGISRLPPEAPPTPTAPPPSPRSEGHDPPT----- 172
| | | | |
DB 115 AEPSTTESADA-----SAVEGDKRAKDEKAAATMLSLQAGRSSSTGPRELKE 167
| | | | |
QY 173 ---ESRSCSQ-----PPSPASDPKACNMKKYKTYLVNSQASQASL--- 212
| | | | |
DB 168 ERGQAESASSGAQETEKADAREPPVELKDPPTS-----SMAAEARALSSES 216
| | | | |
QY 213 ---VGERSSGQPCPQ-----ARLPSEDEASSSSSSSSSESPIPGP 253
| | | | |
DB 217 SEQEMEVBPASKGEQGELEGAGPATVKEGHNHNGEPRENESSAGTDSQELGMEQ 276
| | | | |
QY 254 QSRLEPTAATVQF-----CGAPASTPYLLTSQAODTSGSPSEARPLPGESEF 302
| | | | |
DB 277 NLRSGTGVDRTESKAYGSIHKCEDCGKFT-----HTGNFKRHIRHTGEKPF 325
| | | | |
QY 303 SCQNGEAV---ACSSGLDLSVPDEDEKPYKQLCRSSFRY----- 340
| | | | |
DB 336 SCRECSKAFSDPAACKAHKTHSP---LKPGECECGKSYRLISLNLHKRHSGEARVR 382
| | | | |
QY 341 -----KGNLASHRTVTGGEKYHCISICG----- 363
| | | | |
DB 383 CGDCCGLPTTSNLRQHVHSGQKPYQCDYGRSPDPTSMRRLETHDPTDKHKCHPC 442
| | | | |
QY 364 -ARFNRPANLKTSHRHSGEKPYKCTCGSRFVQVRSPPSGFQGRAGVQKQGFCS 422
| | | | |
DB 443 DKFPNQVGLKAKHLKIHLADGPLKCECKQGT-----TSGNLKRHLRHSGEKPYVCT 496
| | | | |
QY 423 SGRDLKSPSPQVALRAHVRIHTGEKPRPCCTGFRRLQTLKSHVRIHTGEKPYHCD 482
| | | | |
DB 497 HCQRFPADPGA---LGRVRIHTGEKPCQCVICGAFTQASLLAHVQHTGEKPYCE 552
| | | | |
QY 483 PCGLHFRHKSQRRLHR 499
| | | | |
DB 553 RCKGRFVQSSQLANHR 569
| | | | |
RESULT 5
JC7315
C1:Species: Homo sapiens (man)
C1:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 08-Sep-2000
C1:Accession: JC7315
R1:Alliel, P.M.; Seddidi, N.; Goudou, D.; Cifuentes-Diaz, C.; Romero, N.; Velasco, E.; Ri
Biochem. Biophys. Res. Commun. 273, 385-391, 2000
A1:Title: Myonectrin, a novel member of the BTB/POZ-zinc finger family highly expressed in
A1:Reference number: JC7315
A1:Accession: JC7315
A1:Molecule type: mRNA
A1:Residues: 1-610 <ALB>
A1:Cross-references: GB:AF14884
A1:Comment: This protein, belonging to the family of eukaryotic BTB/POZ and zinc finger p
and a regulator in the genital tract.
C1:Keywords: tandem repeat; transcription regulation; zinc finger
Query Match 17.3%; Score 482; DB 2; Length 610;
Best Local Similarity 26.8%; Pred. No. 3.7e-20;
Matches 149; Conservative 65; Mismatches 212; Indels 130; Gaps 16;
QY 16 EFRHSSDVLGNINELRLRGILTDVTLVGGQPLRAHKAVALACGFPYSIFRGRAG--GV 73
| | | | |
DB 2 QVSHHCHEHLRLNQRBAFLCDCTIVIGEFQFAHNVLASFSYFGAIVRSTSENNV 61
| | | | |
QY 74 GVDVLSLGGPRARFAPLIDFMYTSTRRLSPATAPAVLAATYLOMEHVYQACH----- 128
| | | | |
DB 62 FLDQSOV---KADFQKLEBFTYTGTLNLSMWVKEIHQADADYLR/BEVVTYKCKIKMED 117
| | | | |

```

Qy 129 -----RFIOASYEPLGISLRP-----LEABPPTPTAPPG 159
Db 118 FAFIAPSSSTISSITGNIETLNOOTCLTLRDYNNREKSEVSTDLIDANPKQALAKSS 177
Qy 160 -SPRSEGHDPPTESRSCGQ------PSPASBPACMKW---KY 197
Db 178 QTKKKKAFNSPTGQNKTVQYPSDILENASVELFLDANLPPRVVEQVAQINDSELEL 237
Qy 198 KYVLVNSQAG--AGSLVGRSSGQPCQARLPS-----GDEASSSSSSSSSEEPIDP 252
Db 238 TSIVENTFPADIVHTVTVKRGKSGOPNCLKEHSKSNLASKSPLEANSSEE--LDO 295
Qy 253 POSRLSPATATVQFKCAPASTPYLLTSQAQDTSGSPSEARP-----LPSEFFSCQNC 308
Db 296 RYSAKAPMCNT-----CGKVSEASSLRHRIRHKG-----VKPYCHLCGKAFQCNQK 346
Qy 309 AVAGCGSGLDLVPGEDEKRYKQCLCRSSFRYKGNLASHNTVTTG-EKPHGICGARN 367
Db 347 THVRLHTG-----EKPYKCELCDFKQKCOLVPHSRMHGEEKPYKCDVCMQFA 397
Qy 368 RPAHLKTHSRHSGEKPYKCEGSRPVQVRSQPPSGQFKPARGGVGGKGFCSGQROD 427
Db 398 TSSNLKTHAKHSGEKPYKCDRCGRF-----424
Qy 428 LKSPSQVAHLRAHVLHTGEEKPYPCPTGTRFRHLQTLKSHVRIHTEKPYHCDPCGLH 487
Db 425 -----AQASTLTGVHVRHTGEEKPYVCDTGKAFVASSSLITHSRKHTGEEKPYICGICGS 479
Qy 488 FRHKSQRLHLRQKHG 503
Db 480 FISSGLMKHFRSHHG 495

```

RESULT 6

```

A56360
zinc finger protein (clone 18) - human
N/Alternate names: GLI-related finger protein HKR3
C/Species: Homo sapiens (man)
C/Date: 03-Oct-1995 #sequence_reviation 03-Oct-1995 #text_change 05-Nov-1999
C/Accession: A56360; E31201
R/Sugawara, M.; Scholl, T.; Ponath, P.D.; Strominger, J.L.
Mol. Cell. Biol. 14, 8438-8450, 1994
A/Title: A factor that regulates the class II major histocompatibility complex gene DPA
A/Reference number: A56360; MUID:95059073; PMID:7569177
A/Accession: A56360
A/Molecule type: mRNA
A/Residues: 1688 <SUG>
A/Cross-references: GB:U16896; NID:9292934; PIND:AAA65124.1; PID:9292935
R/Ruppert, J.M.; Kinzler, K.W.; Wong, A.J.; Bigner, S.H.; Kao, F.T.; Law, M.L.; Senanez,
Mol. Cell. Biol. 8, 3104-3113, 1988
A/Title: The GLI-Kruppel family of human genes.
A/Reference number: A93103; MUID:89096896; PMID:2850480
A/Accession: E31201
A/Molecule type: DNA
A/Residues: 461-516 <RUP>
C/Genetics:
A/Note: 506/1
A/Note: the list of introns may be incomplete
C/Keywords: DNA binding; zinc finger
F/12-111/Domain: POZ domain homology <POZ>

```

```

Query Match 17.2%; Score 481; DB 2; Length 688;
Best Local Similarity 27.7%; Pred. No. 4.8e-20;
Matches 160; Conservative 72; Mismatches 216; Indels 130; Gaps 22;

Qy 17 FTRHSDVLTGNLRLRGLITVTLVGGQPLRAHKAVALIACSGFYIFRAGAVGD 76
Db 5 FVGHSTRVLTGLNKKQKQYCDATLDVGLVFKAHMSVLACSHFQSLYGGSGGSV- 63
Qy 77 VLSLPGPEKRGAPLIDFMYTSLRLSPATPAVLAATYLTOMEHVQAQCHRT----- 131
Db 64 --VLPAQ-FAEIRGLLDFYTGHLALTSGRDQVLLAAELRVPAVELCQSFKPKTSV 120

```

```

Qy 132 -QAS-----YBPLG-----ISLRPLEAEP--PTPTAPPGSP 161
Db 121 GQAGGQSGSLGPASQNNVSHVKEPAGLEEBEVSRTLGLVPRDQPRGSHSPQRQLHSP 180
Qy 162 RSEGHDPDPT------SRSC-----QGPSAPSPDKACMKWKYITV-- 202
Db 181 AQSEG-----PSSLCKLKOALKPCSLDEKPEDCVPPRPLEBGAOLQGSNEMEVVQ 236
Qy 203 -----NSQASQAGSLVGRSSG---OPCQARLPSGDEASSSSSSSSSEEPIDP 253
Db 237 VEDDGDSTWSEPEAVLTRKSNVTRKPC-----AAEPALASGLAE-----P 280
Qy 254 QSRLSPTATVQFKCAPA-----STPYLLTSQAQDTSGSPSEARPPLGSEFFSCQNC 307
Db 281 AENRGTAVPVE-----CPTCHKKFLSKYLLKVNHRKHGEEKFE--CPKCGCYFRKENTL 334
Qy 308 --EAVAQCGSGLDLVPGEDEKRYKQCLCRSSFRYKGNLASHRTVHTGKPYHICGAK 365
Db 335 LEHEARNCMNRSEQVLP-----CSVQCEFRFRRMELRVHVSHTGEMPYKSSCSQO 386
Qy 366 FMRPANKTHS--RHSGEKPYKCEGSRPV--QVRSQPPSGQFKPARGGVGGKGFCSG 423
Db 387 FMQKQDQSGHMTKLHGAPRPAKPTCAKCFLSRTQLHBAFKHGEKLFVCEGCHRAS 446
Qy 424 QRODLK-----SP------SQVAHLRAHVLHTGEEKPYPCPTGTRFRHLQ 464
Db 447 SRNGLOMHIKAKHNRBRPVHCFCSHAFTQKANLMLHRLTHGKPYFQCHLCKGTFRTQA 506
Qy 465 TLKSHVRIHTEKPYHPCDGLHFRHKSQRLHLRQKH 502
Db 507 SLDKHNRHTHGERPFSCEQRFTEKGPLLHNVASRH 544

```

RESULT 7

```

JH0500
zinc finger protein 29 - mouse
C/Species: Mus musculus (house mouse)
C/Date: 31-Mar-1992 #sequence_reviation 31-Mar-1992 #text_change 08-Dec-2000
C/Accession: JH0500
R/Denny, P.; Ashworth, A.
Gene 106, 221-227, 1991
A/Title: A zinc finger protein-encoding gene expressed in the post-meiotic phase of sper
A/Reference number: JH0500; MUID:92039080; PMID:1937051
A/Accession: JH0500
A/Molecule type: mRNA
A/Residues: 1-614 <DEN>
A/Cross-references: GB:X55126; NID:955470; PIND:CAA38920.1; PID:955471
A/Experimental source: testis
C/Keywords: DNA binding; zinc finger
F/217-244/Region: zinc finger
F/245-272/Region: zinc finger
F/273-300/Region: zinc finger
F/301-328/Region: zinc finger
F/329-356/Region: zinc finger
F/357-384/Region: zinc finger
F/385-412/Region: zinc finger
F/413-440/Region: zinc finger
F/441-468/Region: zinc finger
F/469-496/Region: zinc finger
F/497-524/Region: zinc finger
F/525-552/Region: zinc finger
F/553-580/Region: zinc finger
F/581-608/Region: zinc finger

```

```

Query Match 15.6%; Score 434.5; DB 2; Length 614;
Best Local Similarity 30.3%; Pred. No. 1.8e-17;
Matches 122; Conservative 42; Mismatches 133; Indels 105; Gaps 15;

Qy 162 RSEGHDPDPTESRSCQSPSPASP--DPKACMKWKYITVLSQASQAGSLVGRSS- 218
Db 170 QQSDGDS--FERDCSGGAQGHAPGSDP------VPSBGRVGGILIGTGY 216

```

QY 219 -CG---PCPOA-----RLEPG-----DEASSSSSSSSSSSEEGPIPGQ 254
 DB 217 LGEKPEPCQCKTFSRKSHLITHERHTGKEYKDECKGKSFSDSNFSRHQTHTGK 276
 QY 255 SRLSPAAVQPK---CGAPAS--TPYLLISQAODTSGSPSERA-----R 294
 DB 277 -----PYKCRDCGKSFSSRSANLITHQRIHTGKEKFPQCECKGKSFSSPNLIAHQ 326
 QY 295 PLPGSEFFSCQNCENAVAGSSGIDS---LVPGDEDPYKCOLCRSFRYKGNLASHRTYH 351
 DB 327 THTGKEPYSCPECGKSFGRSSILNTHQGIHTG--EKPYAKCKGEGSFYNSNLIRHQRH 384
 QY 352 TGEKPYHCSICGARFRRPANKLTHSHIHGKEPYKCTGCSRFVQYRSQPSGFGKPAR 411
 DB 385 TGEKPYKCTECGQKFSQSSALITHRTHTGKEPYQCEGCKNF----- 427
 QY 412 GGVGQKGFSSQRODLKSP-----SQVAILRAVLIHTGKEPYPCPTGTRR 461
 DB 428 ---SSNSNLATRRHTLVKPKYKGLGKGSFSSQSSSLIAQGTHTGKEPYECLTGCESES 484
 QY 462 HLQTKSHVRIHTGKEPYHCDPCGLHFRHKSQRLRLRQXG 503
 DB 485 WSNNLKHKRTHTGKEPYKCGDCGKGFSGRSQSLVHQRTHTG 526

RESULT 8

153869
 zinc finger protein - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 01-Dec-2000
 C:Accession: 153869
 R:Brady, J.P.; Platiogoraky, J.
 Gene 149, 299-304, 1994
 A:Title: A mouse cDNA encoding a protein with zinc-fingers and a KRAB domain shows siml
 A:Reference number: 153869; MUID:95047492; PMID:7959006
 A:Accession: 153869
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Accession: 153869
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-555 <RES>
 A:Cross-references: GB:I28167; NID:9758660; PIDN:AAA67545.1; PID:9758661
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 15.5%; Score 432.5; DB 2; Length 555;
 Best Local Similarity 28.6%; Pred. No. 2.1e-17;
 Matches 135; Conservative 52; Mismatches 178; Indels 107; Gaps 18;
 QY 79 SLPGCBPAGPAPLDFMTSLRLSPATAPAVLAATTIQEHVYQACHRFIOASYEPL 138
 DB 112 SYVNTPTQLGKAP-----CQSRPRSPAPVSDENYIK-SHTISV----- 149
 QY 139 GISLRPLEAPPTPTAPPSPG-----PRRSEGHDP-----PP 171
 DB 150 -----PIHSGRPPLPLRAQGTSLTRQTRKQOYHOGSHFDEAGQKRSHTSWIPRR 204
 QY 172 TESRSG-SQGPSPSPAS---PDPKACMKYKYIVILNSQAEQAGSLVGERSSGQPCPQAL 227
 DB 205 SDHRCVCKSGKCSPPDHGRDGVVFPDQNGGIHTQKXSQREAVATDLSSSP--PHQL 262
 QY 228 PSQDEASSSSSSSSSS-----SSEEGPIPGCPSRLSPAAVQFKCGAPAST 274
 DB 263 DSGGRCACAVCECGRPGCTGDCIHOGGRGTGERKEFGKRVCTPKPYTCEQC-----E 317
 QY 275 PYLLTSGAODTSGSPSERARPLPGSEFFSCQNCENAVAGSSGIDS---LVPGDEDPYK 331
 DB 318 KSLVDQHLMSHVKVTTRRP-----YNCERTGSAFSQASHLOHQLHTG--EKPKK 369
 QY 332 QLCRSSFRYKGNLASHRTYHTGKEPYHCSI CGARFRRPANKLTHSHIHGKEPYKCTG 391
 DB 370 DACGKSFSSSHLRSHQVHTGKEPYKCGEGKSFCSNLSLYIHQHVHTGKEPYKCVDDG 429
 QY 392 SRFVQYRSQPSGFGKPARGV--GQXGFGS--SQRODLKSPQVAILRAVLIHTG 447
 DB 430 KEFSR-----PSLQ--AHQGIHTGKSYVCTMCKGYTLNS-----NLQVHLRVHTG 475

QY 448 EKPYPCPTGTRFRHLQTLKSHVRIHTGKEPYHCDPCGLHFRHKSQRLRLH 499
 DB 476 EKPYSCDVQKGFSSRSQSLQSHQVHTGKEPYKCDVCKGKSGWRNSNLIIHR 527

RESULT 9

G02075
 transcription repressor zinc finger protein 85 - human
 C:Species: Homo sapiens (man)
 C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 01-Dec-2000
 C:Accession: G02075
 R:Poncelet, D.A.
 submitted to the EMBL Data Library, September 1995
 A:Reference number: G09169
 A:Accession: G02075
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-595 <PON>
 A:Cross-references: EMBL:U35376; NID:g1017721; PIDN:AAA79179.1; PID:g1017722
 C:Genetic: GDB:ZNF85
 A:Gene: GDB:ZNF85
 A:Cross-references: GDB:132279
 A:Map position: 19p12-19p12
 C:Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 15.2%; Score 424; DB 2; Length 595;
 Best Local Similarity 38.5%; Pred. No. 6.7e-17;
 Matches 90; Conservative 26; Mismatches 92; Indels 26; Gaps 5;

QY 267 KCGAPASTPYLLTSGAODTSGSPSERARPLPGSEFFSCQNCENAVAGSSGL---DSLVRG 323
 DB 206 ECG-----KAFWSSSTLTGKRIHTGKEPYKCECGKAFNQSNNLIKHKIHTG 254
 QY 324 DEDKPYKCOLCRSSFRYKGNLASHRTYHTGKEPYHCSI CGARFRRPANKLTHSHIHG 383
 DB 255 --EKPYKCECGKAFNRPRSTLTTHKIHTGKEPYKCGKAFNRSSTLTTHRKIHTGK 312
 QY 384 PYKCEGSRFVQYRSQPSGFGKPARGVGQKGFSSQRODLKSPQVAILRAVLI 443
 DB 313 PYKCEGKAFKQSSNLTTHKI-----HTGKEPYKCKKCGKAP---NSAHILTHTEV 362
 QY 444 IHTGKEPYPCPTGTRFRHLQTLKSHVRIHTGKEPYHCDPCGLHFRHKSQRLH 497
 DB 363 IHTGKEPYKCEGKAFNHSHTLTTHKIHTGKEPYKCKEKGKAFNHSSTLTTHK 416

RESULT 10

S36336
 probable transcription factor PLZF - human
 C:Species: Homo sapiens (man)
 C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
 C:Accession: S36336; S31989
 R:Chen, Z.; Brand, N.J.; Chen, A.; Chen, S.J.; Tong, J.H.; Wang, Z.Y.; Waxman, S.; Zeller
 EMO J. 12, 1161-1167, 1993
 A:Title: Fusion between a novel Krueppel-like zinc finger gene and the retinoic acid rec
 A:Reference number: S36336; MUID:93209216; PMID:8384553
 A:Accession: S36336
 A:Molecule type: mRNA
 A:Residues: 1-673 <CHE>
 A:Cross-references: EMBL:Z19002; NID:g38517; PIDN:CAA79489.1; PID:g38518
 C:Genetic:
 A:Gene: PLZF
 C:Superfamily: POZ domain homology
 C:Keywords: zinc finger
 F:20-118/Domain: POZ domain homology <POZ>

Query Match 15.1%; Score 421.5; DB 2; Length 673;
 Best Local Similarity 25.3%; Pred. No. 1.1e-16;
 Matches 155; Conservative 63; Mismatches 236; Indels 159; Gaps 19;

QY 20 HSDVILGNLNLRLGILTDVTLVGGPRLRAKAVLIACSGFYFYSIF-RGDAVGVDVL 78
 DB 155: Conservative 63; Mismatches 236; Indels 159; Gaps 19;

Db 16 HPTGLCKANQMBRLAGTLCVIVIMVDSQEFHARTVLACTSKMFEILFHRNSQHTYLLDF 75
 QY 79 SLRGEARGFARLDBPMYTSRLSLSPATAPAVLAATYQMEHVQACGRF---TQASV 135
 Db 76 S-----PKTFQOILEYATTLQAKAEDLDLLVABEILEIYELBEOCKLMETIQASD 129
 QY 136 E----- 136
 Db 130 DNDTEATMADGAEEDERKARYLKNIFISKHSSESGVAVAGSLPGPMVDQSPSVST 189
 QY 137 -----PLGISLRPLEAEPPPTPTAPPGRSRBEGHDPPTESRSCS 178
 Db 190 SFGLSAMSPYKPAVDSIMTIGOSILOTLQPPAGPEEPPLAGGRRPGVAEVTEMQOYD 249
 QY 179 QGPPSPASPDPRKACNMKKYKYIVLNSQASQAGSLVGRSSGQPCPOAR-----LPS 229
 Db 250 E---VBSQDSFGAAE-----SLISGKMGOKVEBKEGEGTPTRSSVITSARELHY 297
 QY 230 GDEASS-----SSSSSSSSSEEGPIPGPQSRSLSPATVQPKCGAPASTPYLLTS--- 280
 Db 298 GRESABQVPRPAADQARTGRRENHAPREPKLGIYSVLPMKADAVLSMPSVTSGLH 357
 QY 281 --QAQTSQSPSRAPRLPGS---EPFSCQCEAVAGCGSLDVLVPEDEKPYKQQLC 334
 Db 358 VQPALAVSMDFTSYGGILLPGQFLQRELFSS-KLGEVLAVGMS--ESRTTIGQ-----CSYC 409
 QY 335 RSSEFRYKGNLASRHYTHTEKPHGSGICGARFMRPNLKTSHRIHS-GEKPYKCEGCGSR 393
 Db 410 GVLPDNEAVEQIRKHLHSGMKTYGCELCGRFLDSLRLNHLHLAHGAKAFCDQCGAO 469
 QY 394 F-----VQVRSQPPSG-----FQGRPA-----RGVGQKGGFCSSQRODLKS 430
 Db 470 FSKEDALETHROHTHTGDMAVFCILCGKRFQAOQALQOHMEVHAGVRSYICSECNRTF-- 527
 QY 431 PRQVAVHLRAVLIHTGEKPYPCPTGTRRHQLTKSHRITITGEKPYKCDPCGGLHFRH 490
 Db 528 -PHTA-LKRLHLSHTGDHPECEFCGSCPRDSTLKSHRITITGEKPYKCNCDKXFLSL 585
 QY 491 KSQLRLHLRQKHG 503
 Db 586 KQOLETHRYVHTG 598

RESULT 11

gammaFAP-C - chicken (fragment)
 C/Species: Gallus gallus (chicken)
 C/Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 01-Dec-2000
 C/Accession: I50643; I50641; I50642; S44242; S44264; S44265
 R/Liu, Q.; Shalaby, F.; Puri, M.C.; Tang, S.; Breitman, M.L.
 Dev. Biol. 165, 165-177, 1994
 A/Title: Novel zinc finger proteins that interact with the mouse gamma F-crystallin prom
 A/Reference number: I50641; MIMD:94374555; PMID:8088434
 A/Accession: I50643
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-676 <LIU>
 A/Cross-references: EMBL:X79011; NID:g475903; PIDN:CAAS5644.1; PID:g475904
 A/Accession: I50641
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 'MRVHRELGMVA', 23, 'GSGR', 28-676 <LI2>
 A/Cross-references: EMBL:X79051; NID:g479055; PIDN:CAAS5653.1; PID:g479056
 A/Accession: I50642
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 36-676 <LI3>
 A/Cross-references: EMBL:X79050; NID:g479057; PIDN:CAAS5652.1; PID:g479058
 C/Superfamily: POZ domain homology
 F:49-152/Domain: POZ domain homology <POZ>

Query Match 15.0%; Score 419; DB 2; Length 676;
 Best Local Similarity 23.7%; Pred. No. 1.5e-16;

Matches 146; Conservative 67; Mismatches 184; Indels 219; Gaps 21;
 QY 9 GALTGYREPTRSSVLTGNLNLRLGILTDVTLVGGQPLRAHRAVLJACGFPYSIR 68
 Db 34 GMLBAMEVPSRQULLQNTQRTKGFCDVITYVQNALFRAHKNITLAASAVLKSLV 92
 QY 69 GRAGVVDVLSLPGBEARG-FAPLLDFMYTSRL-RLSPA---TAPAVLAATYQMEHV 123
 Db 93 -----VHNDLNLNDHEVNSPGIFRLILDFITYRLECEPCEGSGSLGAVLAASVYQIPQL 148
 QY 124 VQACNR-----FIQASYE-PLGISLRPLE----- 146
 Db 149 VALCKKRLKRSKGYCHLRGAPRYKLGRGLRATTPVIOACYSGTRPVDLOVEBAAPLN 208
 QY 147 -----A 147
 Db 209 TCGGELVYASAGTFLHPHGLCPREHSCSPGGLDLSKSPGPSAQLLPTRLRLPAER 268
 QY 148 EPTTTP-----TAPPGSPRRSEGH---DPTESRSCSGPPSP 184
 Db 269 EPLSLPRHDSPPVSGGLAHGPAAYKDSPPGEF---GSHPATDPPSTPPCAE-PLPL 324
 QY 185 ASDPDKAC-NMKKYK---YIVLSQASQAGSLVGRSSGQPCPOARLPS-GDEASS 235
 Db 325 RGDGRELMYRMWKHEPLGPLYL---DEGAKEKLEEREKAESEPPAAPQPPYPSVESNDLEP 381
 QY 236 SSSSSSSSSSEEGPIPG-----POSRLSPATVQPKCGAPASTPYLLTSQA 282
 Db 382 DNSTSEBTSSBEPFGDALDRIYCNHLYGEPSSLGNLYVCIPLCGKFPSSQLAHAYEA 441
 QY 283 QDTs-----GSPERARPLPGESEFSCQCEAVAGCGSLDVLVPEDEKPYKQQLCRSF 338
 Db 442 HNEBELYHKAALAEQVLPF-----LDKGAAGL-----GDILRPYCSGCDKSY 483
 QY 339 RYKGNLASRHYTHTEKPHGSGICGARFMRPNLKTSHRIHSGEKPYPKCFVQVR 398
 Db 484 KDPATLRQEKTHMLTRPYPCTICGKFTQRTWTRHRSHTGLRPFACDAGMRF---- 539
 QY 399 SQPPSGQGPARGVGQGGFCSSQRODLKSPRQVAVHLRAVLIHTGEKPYPCPTG 458
 Db 540 -----TRQRLTBRHRIHSGEKEPBCQVCGG 565

RESULT 12

T14757
 hypothetical protein DKFZp572C163.1 - human (fragment)
 C/Species: Homo sapiens (man)
 C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 01-Dec-2000
 C/Accession: T14757
 R/Wamput, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
 submitted to the Protein Sequence Database, August 1999
 A/Reference number: Z18181
 A/Accession: T14757
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-701 <WAM>
 A/Cross-references: EMBL:AL10217
 C/Genetics: A:Experimental source: adult subthalamic nucleus; clone DKFZp572C163
 A/Note: DKFZp572C163.1
 C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 15.0%; Score 417.5; DB 2; Length 701;
 Best Local Similarity 41.6%; Pred. No. 1.8e-16;
 Matches 89; Conservative 23; Mismatches 77; Indels 25; Gaps 5;
 QY 298 GSEFFSCQCEAVAGCGSGL---DSLVPDEDEKPYKQQLCRSSFPYKGNLASRHYTHGE 354
 Db 352 GKPYECNECEKPTFANNSALRAHONIHG--EKLVECSGCGKTFQKTRLSTHRIRIHGE 409

QY 355 KPYHCSIGCARFNRPNALHTSHRHSIGSEKPYKCECTGSRFVQVRSPSGFGQKPA---- 410
 DB 410 KVEYCSKCGKTFQSKSYLSGHERIHTGEKPYECNCGKTFVY-----KALILYH 458
 QY 411 -RGVQGGKGFSSGSRQDLSKPPSOVAHRAVLIHTGEKPYPCPTGTRFRLIQLTKSH 469
 DB 459 QIHTGEKPYECNCGKTFVY-----SQRTHLCAHQRHTHTGEKPYECNCGKTFADNSALRAH 514
 QY 470 VRIHTGEKPYPCDPCGLHFRHKSQRLHRLRKHG 503
 DB 515 HRIHTGEKPYECNCGKTFQSKSYLSGHERIHTGEKPYECNCGKTFADNSALRAH 548

RESULT 13

zinc finger protein ZNF43 - human
 S26823
 N/Alternate names: zinc finger protein kox27
 C/Species: Homo sapiens (man)
 C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 01-Dec-2000
 C/Accession: S26823; 137967; 510416
 R/Lovering, R.; Trowsdale, J.
 Nucleic Acids Res. 19, 2921-2928, 1991
 A/Title: A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell 11
 A/Reference number: S26823; MUID:91279444; PMID:1711675
 A/Accession: S26823
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-803 <LOV-
 A/Cross-references: EMBL:X59244; NID:g38031; PIDN:CAA11932.1; PID:g38032
 R/Thiesen, H.J.
 New Biol. 2, 363-374, 1990
 A/Title: Multiple genes encoding zinc finger domains are expressed in human T cells.
 A/Reference number: 137949; MUID:91145339; PMID:2288909
 A/Accession: 137967
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 476-531 <THI>
 A/Cross-references: EMBL:X52358; NID:g34160; PIDN:CAA36584.1; PID:g930090
 C/Genetics:
 A/Gene: GDB:ZNF43; HTR6
 A/Cross-references: GDB:128653
 A/Map position: 19p13.1-19p12
 C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 C/Keywords: DNA binding; zinc finger

Query Match 14.9%; Score 415; DB 2; Length 803;
 Best Local Similarity 30.2%; Pred. No. 2, 9e-16;
 Matches 91; Conservative 34; Mismatches 118; Indels 58; Gaps 6;
 QY 267 KCGAPASTVYLLTSAQDNTSGSP-----SERARPLPGSEPFSSQNC-E 308
 DB 312 ECGKAFNWTSTLTGKRHTHTGEKPYCEECGKAFNFSNLTTHKRIHTBKFYKCTECSE 371
 QY 309 AVAGCSGGLDLSVPGDEDEKPYKQLCRSSFRYKGNLASHRTVHTGEKPYHCSIGCARFNR 368
 DB 372 AVSRSSNLTGKHKHTHTGEKPYKCEECGKAFKMSKLTETKHTHTGEKPYKCEECGKAFNW 431
 QY 369 PANLTKHSHRHSIGSEKPYKCECTGSRFVQ-----RSQPPSGFGQKPARGVGQKG 418
 DB 432 PSLTGTGRIHTGEKPYKCEECGKAFNFSNLTTHKRIHTBKFYKCTECSE 489
 QY 419 GFCSSORODLSKPPSOV-----AHLRAVLIHTGEKPYPCPTGTRFRLIQLTKS 468
 DB 490 NLTKHKHTHTGEKPYKCEECGKAFKMSKLTETKHTHTGEKPYKCEECGKAFNFSNLTG 549
 QY 469 HVRHTHTGEKPYHCDPCGLHFRHKSQRLHRLR-----OKHGAATNTVY 510
 DB 550 HKRIHTGEKPYKCEECGKAFNFSNLTTHKRIHTGEKPYKCEECGKAFNFSNLTTHKRI 609
 QY 511 H 511
 DB 610 H 610

RESULT 14

finger protein 1, placental - human
 A32891
 C/Species: Homo sapiens (man)
 C/Date: 20-Oct-1989 #sequence_revision 20-Oct-1989 #text_change 01-Dec-2000
 C/Accession: A32891
 R/Bellefroid, E.J.; Lecocq, P.J.; Benhida, A.; Poncelet, D.A.; Belayew, A.; Martiel, J.J.
 DNA 8, 377-387, 1989
 A/Title: The human genome contains hundreds of genes coding for finger proteins of the
 A/Reference number: A32891; MUID:89377476; PMID:2505992
 A/Accession: A32891
 A/Status: preliminary
 A/Molecule type: mRNA
 A/Residues: 1-428 <BEL>
 A/Cross-references: GB:M27877
 C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology
 C/Keywords: tandem repeat

Query Match 14.8%; Score 413.5; DB 2; Length 428;
 Best Local Similarity 42.0%; Pred. No. 1, 9e-16;
 Matches 87; Conservative 24; Mismatches 77; Indels 19; Gaps 5;

QY 302 FSCQNCBAVAGCSGLDLSVPGDEDEKPYKQLCRSSFRYKGNLASHRTVHTGEKPYH 358
 DB 33 FKCDICGKI FNKKSNTLASHORHTG--EKPYKNCBGKVFHNMSHLAQRRIHTGEKPYK 90
 QY 359 CSICGARFNRPNALHTSHRHSIGSEKPYKCECTGSRFVQVRSPSGFGQKPARGVGQKG 418
 DB 91 CNECGKVFNQSLHQAQRHTHTGEKPYKCEECGKVFHNI-----SHLAQRHTHTGEKPY 144
 QY 419 GFCSSORODLSKPPSOVAHRAVLIHTGEKPYPCPTGTRFRLIQLTKSHVRIHTGEKPY 478
 DB 145 YECNCKGKYF-----SRNSYLVQHLIHTGEKPYKCNCGKVFHNIHSLAQRHTHTGEKPY 200
 QY 479 YHCDPCGLHFRHKSQRL-----RLHLRQK 501
 DB 201 YKNCBGKVFHNMSHLAQRHTHTGEKPYKCEECGKVFHNIHSLAQRHTHTGEKPY 227

RESULT 15

zinc finger protein - human (fragment)
 I37570
 C/Species: Homo sapiens (man)
 C/Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 01-Dec-2000
 C/Accession: I37570
 R/Abirink, M.; Aveskog, M.; Hellman, L.
 DNA Cell Biol. 14, 125-136, 1995
 A/Title: Isolation of cDNA clones for 42 different Kruppel-related zinc finger proteins
 A/Reference number: I37566; MUID:95169271; PMID:7865130
 A/Accession: I37570
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: mRNA
 A/Residues: 1-693 <RES>
 A/Cross-references: EMBL:X78927; NID:g498726; PIDN:CAA5527.1; PID:g498727
 C/Genetics:
 A/Gene: HZF4
 C/Superfamily: zinc finger protein ZFP-36; LIM metal-binding repeat homology

Query Match 14.8%; Score 412.5; DB 2; Length 693;
 Best Local Similarity 41.4%; Pred. No. 3, 5e-16;
 Matches 87; Conservative 18; Mismatches 76; Indels 29; Gaps 4;

QY 298 GSEFSSQNCBAVAGCSGLDLSVPGDEDEKPYKQLCRSSFRYKGNLASHRTVHTGEKPY 356
 DB 419 GSKPYKCEECGKAFNFSNLTTHKRIHTBKFYKCTECSE 478
 QY 357 YHCSIGCARFNRPNALHTSHRHSIGSEKPYKCECTGSRFVQ-----VRSQPPSGFGQ 407
 DB 479 YKCEECGKGFNRRLDKHCRHTHTGEKPYKCEECGKVFHNMSHLAQRHTHTGEKPY 530
 QY 408 KPARGVGQKGFCSSORODLSKPPSOVAHRAVLIHTGEKPYPCPTGTRFRLIQLTK 467

Db	.531	-----GKPKC----	EECGKSFSSAHLQAKVHTGKPYKCGECGKGFKWSLND	579
Qy	468	SHVRIHTGKPYHCDPGLHFRHKSOLRLH	497	
Db	580	MHQRVHTGKPYTCGACGKHFSQASSLQLH	609	

Search completed: July 14, 2003, 18:23:16
Job time : 14.6598 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: July 14, 2003, 18:02:46 ; Search time 5.38782 seconds

(without alignments)
3987.653 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789

Sequence: 1 MGSPAPEGALGYVREFTRH.....ROKGAATVTKVHILGSP 518

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1138	40.8	707	1 BCL6_MOUSE	P41183 mus musculu
2	1128.5	40.5	706	1 BCL6_MOUSE	P41182 homo sapien
3	575.5	20.6	803	1 Z151_HUMAN	Q13105 homo sapien
4	558	20.0	794	1 Z151_MOUSE	Q60821 mus musculu
5	482.5	17.3	697	1 Y441_HUMAN	Q43167 homo sapien
6	476	17.1	688	1 HKR3_HUMAN	P10074 homo sapien
7	434.5	15.6	614	1 ZF28_MOUSE	Q07330 mus musculu
8	427.5	15.3	645	1 ZF93_MOUSE	Q61116 mus musculu
9	424	15.2	595	1 ZN85_HUMAN	Q16500 homo sapien
10	421.5	15.1	458	1 Z239_HUMAN	Q05516 homo sapien
11	421.5	15.1	673	1 Z145_HUMAN	Q14526 homo sapien
12	420.5	15.1	733	1 HIC1_HUMAN	Q96103 homo sapien
13	419	15.0	615	1 HIC2_HUMAN	Q96103 homo sapien
14	419	15.0	676	1 HIC1_CHICK	Q90850 gallus gall
15	418	15.0	892	1 HIC1_MOUSE	Q91715 mus musculu
16	415.5	14.9	803	1 Z226_HUMAN	Q9NYT6 homo sapien
17	415	14.9	476	1 Y106_HUMAN	Q96319 homo sapien
18	415	14.9	803	1 ZN43_HUMAN	P17038 homo sapien
19	413.5	14.8	407	1 OZF_MOUSE	Q62513 mus musculu
20	413.5	14.8	407	1 ZN83_HUMAN	P51522 homo sapien
21	412.5	14.8	698	1 Z234_HUMAN	Q14588 homo sapien
22	412	14.8	597	1 ZP37_HUMAN	Q9Y6G3 homo sapien
23	411.5	14.8	818	1 KR18_HUMAN	Q9HCG1 homo sapien
24	410.5	14.7	606	1 Z214_HUMAN	Q9U159 homo sapien
25	408.5	14.6	616	1 ZN93_HUMAN	P35789 homo sapien
26	407	14.6	209	1 ZF93_HUMAN	Q14590 homo sapien
27	406.5	14.6	574	1 YE73_HUMAN	Q9P255 homo sapien
28	406.5	14.6	629	1 Z195_HUMAN	Q14628 homo sapien
29	404.5	14.5	446	1 ZN38_HUMAN	P17098 homo sapien
30	404	14.5	734	1 ZN42_HUMAN	P28698 homo sapien
31	403	14.4	683	1 Y972_HUMAN	Q9Y2H8 homo sapien
32	402.5	14.4	504	1 Z205_HUMAN	Q95201 homo sapien
33	402.5	14.4	682	1 ZN45_HUMAN	Q02386 homo sapien

34	402	14.4	407	1 OZF_RAT	Q62981 rattus norv
35	401	14.4	706	1 Z151_CHICK	Q90625 gallus gall
36	401	14.4	1191	1 ZN91_HUMAN	Q05481 homo sapien
37	400	14.3	741	1 Z288_HUMAN	Q9HC78 homo sapien
38	399.5	14.3	913	1 Z228_HUMAN	Q9UJ33 homo sapien
39	399	14.3	469	1 Z135_HUMAN	P52742 homo sapien
40	398	14.3	489	1 E21T_HUMAN	Q9NQZ8 homo sapien
41	397	14.2	543	1 ZN08_HUMAN	P17098 homo sapien
42	396.5	14.2	498	1 Z271_HUMAN	Q14591 homo sapien
43	396	14.2	449	1 ZF91_HUMAN	Q43829 homo sapien
44	396	14.2	449	1 ZF91_MOUSE	Q08376 mus musculu
45	396	14.2	604	1 Z300_HUMAN	Q36199 homo sapien

ALIGNMENTS

```

RESULT 1
ID      BCL6_MOUSE          STANDARD;          PRT;          707 AA.
AC      P41183; 061065;
DT      01-FEB-1995 (Rel. 31, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      B-cell lymphoma 6 protein homolog.
GN      BCL6 OR BCL-6.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=BAH/c; TISSUE=Skeletal muscle;
RX      MEDLINE=96038894; PubMed=7478591;
RA      Fukuda T., Miki T., Yoshida T., Hatan M., Ohashi K., Hirosewa S.,
RA      Tokuhisa T.;
RT      "The murine BCL6 gene is induced in activated lymphocytes as an
RT      immediate early gene.";
RL      Oncogene 11:1657-1663 (1995).
[2]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Muscle;
RX      MEDLINE=96247530; PubMed=8652841;
RA      Allman D., Jain A., Dent A., Malle R.R., Selveggi T., Kenny M.R.,
RA      Staedt L.M.;
RT      "BCL-6 expression during B-cell activation.";
RL      Blood 87:5257-5268 (1996).
CC      - FUNCTION: TRANSCRIPTION REGULATOR THAT PROBABLY PLAY AN IMPORTANT
CC      ROLE IN LYMPHOMAGENESIS.
CC      - SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; D38377; BAA07456.1; -
DR      EMBL; U41465; AAB17432.1; -
DR      HSSP; P15822; 1BBO.
DR      MGD; MGI:107187; Bcl6.
DR      InterPro; IPR000210; BTB_POZ.
DR      InterPro; IPR000822; ZnF_C2H2.
DR      Pfam; PF00096; zC2H2; 6.
DR      PRINTS; PR00048; ZINC_FINGER.
DR      SMART; SM00225; BTB; 1.
DR      SMART; SM00355; ZnF_C2H2; 6.
DR      PROSITE; PS50097; BTB; 1.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.

```

DR PROSITE, PSS0157; ZINC FINGER C2H2_2; 6.
 KM Nuclear protein; Transcription regulation; Activator; DNA-binding;
 KW Zinc-finger; Metal-binding; Repeat.
 FT DOMAIN 32 99 BTR.
 FT ZN_FING 519 682 ZINC FINGERS.
 FT ZN_FING 519 542 C2H2-TYPE.
 FT ZN_FING 547 569 C2H2-TYPE.
 FT ZN_FING 575 597 C2H2-TYPE.
 FT ZN_FING 603 625 C2H2-TYPE.
 FT ZN_FING 631 653 C2H2-TYPE.
 FT ZN_FING 659 682 C2H2-TYPE.
 FT CONFLICT 456 456 A -> G (IN REF. 2).
 SQ SEQUENCE 707 AA; 78981 MW; 2051DD808D32D5EC CRC64;
 Query Match 40.84; Score 1138; DB 1; Length 707;
 Best Local Similarity 36.44; Prod. No. 3.1e-51;
 Matches 265; Conservative 51; Mismatches 149; Indels 270; Gaps 14;
 QY 1 MGSPAPREGALGVRETRHSDVLTGNLRLRGILTDVTLVGGQPLRAHRAVLIAQS 60
 1 MASPA--DSCI-----QTRHSDVLTNLRNRSDILTDVIVVSRGQRAHKTIVLMACS 54
 DB 61 GFRYSIFRGAGVGVVLSLPGGPEARGFAPLIDFWTTSRLSPATAPVLAATYLOW 120
 55 GLFYSLFTDQKCNLSYINLDPISBEGFCILDFWYTSRLNLRBNINAVMTTAYLOW 114
 QY 121 EHVYVACHRFIOASYEPLGISLRPLAE----- 148
 115 EHVVDCTCRKITKASEMAMALKPREEPLNMLPHDIMAARGREVENNMLPANTRG 174
 QY 149 -----PPT----- 151
 DB 175 CESRAFAPLYSGLSTPPASYMYSHLPJLSTFLPSDEBLDAPRVANPPEKRALPCD 234
 QY 152 ----- 151
 DB 235 SARQVNEYSRPAMEVSPSLCHSNISPKVAPEARSIDHVSVEGPKPAVPSANAPY 294
 QY 152 -----PPTAP-----PGSPRSEGHDPPTESRS----- 176
 DB 295 FPCDKASKERERSSSEDEILHEPPAPLNRKGLVSPSPQSDQCPNSPTSCSSKQA 354
 QY 177 -----CSQGPSPAPSPPKACNMKKYKIVLNS-----QASQAGSLVGRSSGQPCQA 225
 355 CILQASGSPPAKSPPTPKACNMKKYKIVLNSLNQAKPGSSQALGRLSPRAYAPAPA 414
 DB 226 RLBSGDEAS--SSSSSSSSSSSEEGPPEPQSL-----SPTATVQF----- 266
 415 CQPMPEPANLDLQSPKLSASGEDSTIP--QASRLNMLVNRSLAGSPRSSSHSPMYAP 473
 QY 267 -KCGASAPFPYLLTQAODTSGSPSERARPLPQSEF-----FSCQNCNAVAGCSSG 316
 474 PKTSCGSGSPQHTENCLHTAGTPEEMOETQSEISDSCENGCTPFCNECDRFESEBS 533
 DB 317 LD-SLVGDEDEKPYKQCLRSSFRYKGNLASHRTVATGEKPYKCSICGAFNRPANLKT 375
 534 LKHTLQTHSDKDYKCDRCQASFRYKGNLASHRTVATGEKPYKCNICGAFNRPANLKT 593
 QY 376 SRHSEKPYKCTCGSRFVQVNSQPSGCGKPARAGVCGKGFSSGRODLKSPSQY 435
 594 TRHSEKPYKCTCGARFV----- 615
 DB 436 AHLRAVLLHTGKPYPCPTGTRPHLOTLSKSHVHTHGEKPYHCDPGSLFRHRSQAR 495
 616 AHLRAVLLHTGKPYPCPTGTRPHLOTLSKSHVHTHGEKPYHCEKCNLFRHRSQAR 675
 QY 496 LHLRQKGAATNTKYVHI 514
 676 LHLRQKGAATNTKYVRI 694
 DB

RESULT 2
 BCL6_HUMAN

ID BCL6_HUMAN STANDARD; PRT; 706 AA.
 AC P41B2;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE B-cell lymphoma 6 protein (BCL-6) (Zinc finger protein 51) (LAZ-3 protein) (BCL-5).
 GN BCL6 OR ZNF51 OR LAZ3 OR BCL5.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 NX NCBI_TaxID=9606;
 RX SEQUENCE FROM N.A.
 RC TISSUE=Skeletal muscle;
 RX MEDLINE=94035122; PubMed=8220427;
 RA Kerchaert J.-P., Dewaendt C., Tilly H., Quief S., Lecocq G., Bastard C.
 RT "LAZ3, a novel zinc-finger encoding gene, is disrupted by recurring chromosome 3q27 translocations in human lymphomas."
 RL Nat. Genet. 5:66-70(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94053709; PubMed=8235596;
 RA Ye B.H., Lieta F., Lo Coco F., Knowles D.M., Offit K., Chaganti R.S.K., Dalla-Favera R.
 RT "Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphoma."
 RL Science 262:747-750(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94100541; PubMed=8274740;
 RA Miki T., Kawamata N., Hirosewa S., Aoki N.
 RT "Gene involved in the 3q27 translocation associated with B-cell lymphoma, BCL5, encodes a Kruppel-like zinc-finger protein."
 RL Blood 83:26-32(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93281738; PubMed=8506375;
 RA Barton B.W., Nicotora G., McCabe N., Espinosa R. III, le Beau M.M., McKelthan T.W.,
 RT "Identification of the gene associated with the recurring chromosomal translocations t(3;14) (q27;q32) and t(3;22) (q27;q11) in B-cell lymphomas."
 RL Proc. Natl. Acad. Sci. U.S.A. 90:5262-5266(1993).
 CC -1- FUNCTION: TRANSCRIPTIONAL REGULATOR THAT PROBABLY PLAYS AN IMPORTANT ROLE IN LYMPHOMAGENESIS.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- DISEASE: INVOLVED IN A FORM OF B-CELL NON-HODGKIN LYMPHOMA CHARACTERIZED BY CHROMOSOMAL TRANSLOCATION T(3;14) (Q27;Q32) AND T(3;22) (Q27;Q11) THAT INVOLVES BCL6 AND IMMUNOGLOBULIN GENE REGIONS.
 CC -1- DISEASE: INVOLVED IN A FORM OF B-CELL LEUKEMIA CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(3;11) (Q27;Q23) THAT INVOLVES BCL6 AND POU2AF1 (OBF1).
 CC -1- DISEASE: INVOLVED IN A T(3;4) (Q27;P11) CHROMOSOMAL TRANSLOCATION WITH ARH (TTF).
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.; WWW="http://www.infobiogen.fr/services/chronocancer/Genes/BCL6ID20.htm".
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib.ch).
 CC EMBL; Z21943; CAA79937.1; -;
 CC EMBL; U00115; AAC50054.1; -;
 CC EMBL; S67779; -; NOT_ANNOTATED_CDS.

DR PIR, S32767, S32767.
 DR PIR, S40520, S40520.
 DR HSSP, P15822, 1B80.
 DR TRANSFAC, T02322, -.
 DR Genew; HGNC:1001, BCL6.
 DR MIM, 109565, -.
 DR InterPro, IPR000210, BTB_POZ.
 DR InterPro, IPR000822, ZnF_C2H2.
 DR Pfam, PF00096, zf-C2H2, 6.
 DR Pfam, PF00651, BTB, 1.
 DR PRINTS, PR00048, ZINC_FINGER.
 DR SMART, SM00225, BTB, 1.
 DR SMART, SM00355, ZnF_C2H2, 6.
 DR PROSITE, PS50097, BTB, 1.
 DR PROSITE, PS50028, ZINC_FINGER_C2H2_1, 6.
 DR PROSITE, PS50157, ZINC_FINGER_C2H2_2, 6.
 DR Nuclear protein; Transcription regulation; Activator; DNA-binding;
 KW Zinc-finger; Metal-binding; Repeat; Proto-oncogene;
 KM Chromosomal translocation.
 FT DOMAIN 32 99
 FT ZINC 518 681
 FT ZN_FING 518 681 C2H2-TYPE.
 FT ZN_FING 546 568 C2H2-TYPE.
 FT ZN_FING 574 596 C2H2-TYPE.
 FT ZN_FING 602 624 C2H2-TYPE.
 FT ZN_FING 630 652 C2H2-TYPE.
 FT ZN_FING 658 681 C2H2-TYPE.
 FT ZN_FING 347 347 S -> A (IN REF. 2).
 FT CONFLICT 393 393 E -> G (IN REF. 2).
 FT CONFLICT 498 498 P -> A (IN REF. 3).
 FT SEQUENCE 706 AA; 78846 MW; E38D83C213DAE2D0 CRC64;

Query Match 40.5%; Score 1128.5; DB 1; Length 706;
 Best Local Similarity 35.0%; Pred. No. 9,3e-51;

Matches 270; Conservative 48; Mismatches 117; Indels 337; Gaps 16;

QY 1 MGSPAEGALGVREFTTHSSDVLGNLNLRLGLITDVTLLVGGQPLRAHKAVALIACS 60
 DB 1 MASPA--DSCI---QFTRHASDVLNLNRLRGRDILTDVIVVSEQGRRAKHTVLMACS 54
 QY 61 GFFYSIFRRAAGVGVVLSLPGGPEARGAPLIDFMTYTRLRISPTAAVLAATYLLQM 120
 DB 55 GLFYSIFRRAAGVGVVLSLPGGPEARGAPLIDFMTYTRLRISPTAAVLAATYLLQM 114
 QY 121 EHVVOCHRFIONS----- 134
 DB 115 EHVVDCTKRFKASAEAMVSAIKPRPEEFNLNMLPMODIMAYRGREVENNLPLRSAGC 174
 QY 135 -----YELGIS----- 141
 DB 175 CESRAVAPSLVGLSTPPASVSHLPVSLFSDDEFNDVMPVAPNPKERALPCDS 234
 QY 142 -----LRL----- 145
 DB 235 ARPVGEVAPTLVSPVNVCHSNIVSPKETIPEEARSDMHYVAEGLKPAASARNAVYF 294
 QY 146 -----EABPT-----PPTAP-----PGSPRREEGHPDPTSRSS----- 176
 DB 295 PCDKASKEERPSSEDEILAFEPNAPLNRKGLVSPQSDQCPNPSPTSCSKNAC 354
 QY 177 ---CSGQPPSPAPDPDKACNMWKYKTYIVNSQASQAGSLVGRSSGQPCPQARLPAGDEA 233
 DB 355 ILQAGSPPAKSPPTDPAKCNMKYKTYIVN----- 384
 QY 234 SSSSSSSSSSEEGPIPGQSRSLPTAATVQFKCAP-----ASTYLLTSQAQDT- 285
 DB 385 ---SLNQNAKPEGEQAEELGLSPRAYTAPACQPMIPENLDQSPTKLSASGEDST 439
 QY 286 -----SGSP-----SERAPL----- 296
 DB 440 IQQASHLNNIIVNRSMGTSPSSSESGSLYMHAPKCTSCQSOSQDAEMCNLHTAGTPEE 499
 QY 297 ---PGSER-----FSCNCEAVAGCGSLD-SLVPGDEKPYKCOLCRSPFRYKG 342

DB 500 EMEGTQSEYSDSSCENGAFPCNECDRPSSEASLKRHTLQTHSDAPRYKDRQASFRYKG 559
 QY 343 NLSHRTVHTGKPYHCISGARFNRPNANLKTSHRISHGKPYKCTGSRPVQYRSQPP 402
 DB 560 NLSHRTVHTGKPYHCISGARFNRPNANLKTSHRISHGKPYKCTGSRPVQYRSQPP 612
 QY 403 SGFOGKPARGVGVQKGFCSQRODLKSPSQVAHLRAHVLHTGKPYKCTGSRPRH 462
 DB 613 -----QVALRLRAHVLHTGKPYKCTGSRPRH 641
 QY 463 LQTLKSHRHTGKPYHCISGARFNRPNANLKTSHRISHGKPYKCTGSRPVQYRSQPP 514
 DB 642 LQTLKSHRHTGKPYHCISGARFNRPNANLKTSHRISHGKPYKCTGSRPVQYRSQPP 693
 RESULT 3
 Z151_HUMAN STANDARD; PRT; 803 AA.
 AC Q13105; Q15932; Q9NDC9;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Zinc finger protein 151 (Myc-interacting zinc finger protein) (Miz-1 protein).
 GN ZNF151 OR MIZ1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97453565; PubMed=9308237;
 RA Schneider A., Peukert K., Eilers M., Haenel F.;
 RT "Association of Myc with the zinc-finger protein Miz-1 defines a novel pathway for gene regulation by Myc.";
 RT Curr. Top. Microbiol. Immunol. 224:137-146(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Bird C.;
 RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 580-803 FROM N.A.
 RC TISSUE=Insulinoma;
 RX MEDLINE=96044430; PubMed=7557990;
 RA Tommerup N., Vissing H.;
 RT "Isolation and fine mapping of 16 novel human zinc finger-encoding cDNAs identify putative candidate genes for developmental and malignant disorders.";
 RT Genomics 27:259-264(1995).
 RN [4]
 RP SEQUENCE OF 327-342 FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=92372070; PubMed=1505991;
 RA Lichter P., Bray P., Ried T., David I.B., Ward D.C.;
 RT "Clustering of C2-H2 zinc finger motif sequences within telomeric and fragile site regions of human chromosomes.";
 RT Genomics 13:999-1007(1992).
 CC -1- SUBUNIT: ASSOCIATED WITH THE C-TERMINAL OF MYC.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/announce/> or send an email to license@ebi.ac.uk).
 CC EMBL; U20647; AAC50256.1; -.


```

KW Nuclear protein; Repeat.
PT DOMAIN 1 104 BTB. FINGERS.
PT ZINC 297 730
PT ZN FING 297 319 C2H2-TYPE.
PT ZN FING 325 347 C2H2-TYPE.
PT ZN FING 353 375 C2H2-TYPE.
PT ZN FING 381 403 C2H2-TYPE.
PT ZN FING 409 431 C2H2-TYPE.
PT ZN FING 437 459 C2H2-TYPE.
PT ZN FING 465 487 C2H2-TYPE.
PT ZN FING 493 515 C2H2-TYPE.
PT ZN FING 519 543 C2H2-TYPE.
PT ZN FING 549 571 C2H2-TYPE.
PT ZN FING 577 599 C2H2-TYPE.
PT ZN FING 605 628 C2H2-TYPE.
PT ZN FING 708 730 C2H2-TYPE.
PT ZN FING 507 527 G -> A (IN REF. 2).
PT CONFLICT 573 573 N -> K (IN REF. 2).
SQ SEQUENCE 794 AA; 86664 MW; FF8856DEBFD7ED CRC64;

Query Match
Best Local Similarity 20.0%; Score 558; DB 1; Length 794;
Matches 164; Conservative 72; Mismatches 199; Indels 182; Gaps 17;

QY 16 EFTRHSDVLYGNLNLRLGLITDVTLLVGGQPLRAHKAVLIACSGFFYSIFRGRAG-VG 74
DB 2 DEFGHSQRYLEQDINQORQLGLDCDCTFVVVDGVDFAHKAVALAACSEYFKMLFVDQKDVH 61
QY 75 VVVLISLPGPEARGAPLIDPMYTSRLRSPATAPAVLAATYLOMEHNVQACHRTQAS 134
DB 62 LDI-----SNAGLGQVLEFMYTALSLSPENVDDVLAVASFLQMDIVTACTH-LKSL 114
QY 135 YEPPLGISLRLPAEPPTPTAPPSPRSSEGHDPPT----- 172
DB 115 AEPSSITGSSAA-----SAVEGDKAKADEKAATMLSRUGARGSSITGPELKE 167
QY 173 ---BSRSCSQG-----PPSPAPDPKAKCMKKYKYYIVLNSQASQAGSL--- 212
DB 168 ERGQASASGAEQTEKADAPREPVELKDPPTS-----SMAAFAEALSSS 216
QY 213 ---VGRSSSGQPPQ-----ARLPBGDDASSSSSSSSSSEEGTIPRP 253
DB 217 SFOEMEVEPASGEGDQEEGAGPATVKEGMLDNGEPREMEESAGTDSGQELMEQ 276
QY 254 QSRLSEPTATVQFK-----CGAPASFPYLLTSQMDTSGSPSRARPLPGESEF 302
DB 277 NLRSGTYGRTESKAVGSIHKCEDCGKEPT-----HTGNFKAHIRIHTEKEFP 325
QY 303 SCONCEAV-----AGSSGGLDSLVPGDEDPKQYKQLCRSSPRY----- 340
DB 326 GCRGEGSKAPSDPAACAHKHTSP---LKPYGCEGCKGKRYLLSLNLHKHSGBARVR 382
QY 341 -----KGNLASHRTVHTGEKRYKHSICG----- 363
DB 383 CGDCCGLPTTSGNLKHLQVHSGQKYOCDYCRSPSPDPSKMRHLEHTDTEHNCPRHC 442
QY 364 -ARPNRPATLKTHSRHSGEKPKCECTGSRFPVQVNSOPSPSGOGKPARGVGQKGFSS 422
DB 443 DKKPNQVGLKAKHLKHLIDGPKCKCEGCKQFT-----TSGNLKHLHISGEKPYVLT 496
QY 423 SGRODLKSPSOVAHLRAHVLHTGEEKPYPCPTGTRFRHLQTLKSHVHHTGEKEPYND 482
DB 497 HGORQPADP-----GGLGRVYRIRHTGKPCQCVLCGAFLYQASLIHVNQHTGEKPYVE 552
QY 483 PCGLHFRHNSQLRHLNR 499
DB 553 RCGKRPVQSSQLANHR 569

```

```

DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical zinc finger protein KIA00441.
GN KIA00441.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=brain;
RC MEDLINE=98116655; PubMed=9455477;
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VIII.
RT 78 new cDNA clones from brain which code for large proteins in
RT vitro."
RL DNA Res. 4:307-313 (1997).
CC -1- FUNCTION: MAY BE INVOLVED IN TRANSCRIPTIONAL REGULATION.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB007901; BAA23713.1; -.
DR HSSP; P08046; 1AIG.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR000210; BTB POZ.
DR InterPro; IPR000822; ZnF C2H2.
DR Pfam; PF00096; zf-C2H2; 8.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF02178; AT hook; 1.
DR PRINTS; PR00048; ZINC-FINGER.
DR ProDom; PD000003; ZnF C2H2; 2.
DR SMART; SM00384; AT hook; 1.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00355; ZnF C2H2; 8.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Zinc-finger; Metal-binding; Nuclear protein; Repeat.
PT DOMAIN 10 133 BTB.
PT ZINC 294 512 ZINC FINGERS.
PT ZN FING 294 316 C2H2-TYPE.
PT ZN FING 322 344 C2H2-TYPE.
PT ZN FING 350 372 C2H2-TYPE.
PT ZN FING 378 400 C2H2-TYPE.
PT ZN FING 406 428 C2H2-TYPE.
PT ZN FING 434 456 C2H2-TYPE.
PT ZN FING 462 484 C2H2-TYPE.
PT ZN FING 490 512 C2H2-TYPE.
SQ SEQUENCE 697 AA; 78292 MW; F2BD33C144626544 CRC64;

Query Match
Best Local Similarity 17.3%; Score 482.5; DB 1; Length 697;
Matches 156; Conservative 74; Mismatches 225; Indels 133; Gaps 19;

QY 6 APEGALGYVREFTRHSDVLYGNLNLRLGLITDVTLLVGGQPLRAHKAVALIACSGFFYS 65
DB 5 SPSPGQLVHSDAHSVDVLAASFEDQKKGFCDITLVENVHPRDRLALAASSSEYFEM 64
QY 66 IFRGRAGVQVDVLSLPGPEARGAPLIDPMYTSRLRSPATAPAVLAATYLOMEHNVQ 125

```


DB 65 MFAEBEIGQSIYMLE-GWVADTFGLLEFYLYGLHASEKSTEQIILATQAFUKVYDLVK 123
 QY 126 ACHRFIOASYEPGLISLRLEAPPTPTAPP--GSPRSEGHPPPTESRESCGSPS 183
 DB 124 AYIDFQNNHSP-----KPTLTNAGAVVIVIPNKA--DPEKPKG----- 163
 QY 184 PASPDKACNMWKYK-----YIVNSQASQASGLVGERSSQCPQOARLPBGDEA 233
 DB 164 ----RKKVNTLQEEKSELAAEBEIDLRVNSVQNRQNFVVKDGSVLNBOI---AAEK 216
 QY 234 SSSSSSSSSSSSEGP1-----PGQSLSPFA--ATYQFCGABASTPYLL 278
 DB 217 ESEEPFCEPSREEMVEKEDENYDKPTEDQASQSRYSKRRIRWRSYVLK-----DYKL 269
 QY 279 TSOADPTSGSPSE--DARLPGESEFPSCNCEAV-----ACSSGLDLVPGDD 326
 DB 270 VGDQED-KGSAKKICGRKRRKPGPEARCCDQGVFKYNNPLAHQSHTG-----E 319
 QY 327 KPYKCOLCRSSFFRYKNLASHRTVHTGEEKYHCSIGARFNRPAULKTHSRHSGEKP 386
 DB 320 RPKKNECGKGFAGKSLQVHTMHTGERPYCTVCSKALTGSLLEHMSLHSGKSP 379
 QY 387 CETSGRFQVNRQPSGPGKPARAGVGOKGFCSSORODLSPSQVAHLRAVYLIT 446
 DB 380 CDQCGRYFQNR-QLKSHYR-----VHTGSHLPECKDCHRKEMDVSLKGLRTHT 429
 QY 447 GEKPYCPGCTGTFRLQTLKSHVRLHTGKPYHCDPCGLHF-----RH----- 490
 DB 430 GEKPYFCEICGKSFYKASSLQTHIRIRHGEKPSGICGKSPSDSAAKRRHCLHTGK 489
 QY 491 -----KSLRLRLRQKGA-----ATNTKVHYHL 515
 DB 490 FSCPECNLQFARLDNKAHLKHSKEMASDASIGSSNTEVERNIL 537

RESULT 6
 HKR3 HUMAN STANDARD; PRT; 688 AA.
 ID P10074;
 AC 01-MAR-1989 (Rel. 10, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 GN Krueppel-related zinc finger protein 3 (HKR3 protein).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95059073; Pubmed=7969177;
 RA Sugawara M., Scholl T., Ponath P.D., Stronginger J.L.,
 "A factor that regulates the class II major histocompatibility
 complex gene DPA is a member of a subfamily of zinc finger proteins
 that includes a Drosophila developmental control protein.",
 RT Mol. Cell. Biol. 14:8438-8450 (1994).
 RL (2)
 RN SEQUENCE FROM N.A.
 RP MEDLINE=98177483; Pubmed=9516840;
 RA Maris J.M., Jensen J., Sulman E.P., Beltlinger C.P., Allen C.,
 Biegel J.A., Brodeur G.M., White P.S.,
 "Human Kruppel-related 3 (HKR3): a candidate for the 1p36
 neuroblastoma tumour suppressor gene?",
 RT Eur. J. Cancer 33:1991-1996 (1997).
 RN (3)
 RP SEQUENCE OF 461-488 FROM N.A.
 RA MEDLINE=89096896; Pubmed=2850480;
 RA Ruppert J.M., Kinzler K.W., Wong A.J., Bigner S.H., Kao F.T.,
 Law M.L., Suarez H.N., O'Brien S.J., Vogelstein B.,
 "The Gli-Kruppel family of human genes",
 RT Mol. Cell. Biol. 8:3104-3113 (1988).
 CC -1- FUNCTION: BINDS TO AND REGULATES THE J AND/OR S ELEMENTS IN MHC II
 CC PROMOTER.

CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 CC EMBL, L16896; AAA65124.1; .
 CC EMBL, U45325; AAB08973.1; .
 CC EMBL, U45324; AAB08973.1; JOINED.
 CC EMBL, M20677; AAA35989.1; .
 CC PIR, E31201; E31201.
 CC HSSP, P03001; 1TF3.
 CC TRANSFAC, T04971; .
 CC Genew; HGNC:4930; HKR3.
 CC MIM: 165270; .
 CC
 CC InterPro; IPR000210; BTB_POZ.
 CC InterPro; IPR000822; Znf_C2H2.
 CC Pfam; PF00096; ZF-C2H2_11.
 CC Pfam; PF00651; BTB; 1.
 CC PRINTS; PR00048; ZINC_FINGER.
 CC ProDom; PD000003; Znf_C2H2; 1.
 CC SMART; SM00225; BTB; 1.
 CC SMART; SM00355; Znf_C2H2; 11.
 CC PROSITE; PS50097; BTB; 1.
 CC PROSITE; PS50028; ZINC_FINGER_C2H2_1; 9.
 CC PROSITE; PS50157; ZINC_FINGER_C2H2_2; 11.
 CC Transcription regulation; Activator; Zinc-finger; DNA-binding;
 CC Repeat; Metal-binding; Nuclear protein.
 CC
 CC BTB
 CC DOMAIN 26 89
 CC FT 291 600
 CC FT ZN FING 291 313
 CC FT ZN FING 319 342
 CC FT ZN FING 350 372
 CC FT ZN FING 378 401
 CC FT ZN FING 407 430
 CC FT ZN FING 436 459
 CC FT ZN FING 465 487
 CC FT ZN FING 493 515
 CC FT ZN FING 521 544
 CC FT ZN FING 550 572
 CC FT ZN FING 578 600
 CC FT CONFLICT 201 201
 CC FT CONFLICT 244 244
 CC FT CONFLICT 350 351
 CC FT CONFLICT 607 607
 CC FT SEQUENCE 688 AA, 77054 MW; EBECCB3DC6BDB524 CR64;
 CC
 CC Query Match 17.1%; Score 476; DB 1; Length 688;
 CC Best Local Similarity 27.1%; Pred. No. 1,1e-17;
 CC Matches 157; Conservative 73; Mismatches 216; Indels 134; Gaps 22;
 CC
 CC 17 FTRHSSDVLGNINERLRLGILTDVTLVGGOPLRAHKAVALVLCGFFPISFGRAGVCD 76
 CC 5 FQOHSVRYLQELNKKREKQYCDATLDVGLVFKAWWSVLACSHFPGSLVYDGGSGSV- 63
 CC 77 VLSLGGPEARGFALDPMYTSRLSPATPAVLAATVYQMHVVOACGRFI----- 131
 CC 64 --VLPAG-FAELFGLLDFTTGHALTSNGRDVLLARELRVEANVELCSFKPKTSV 120
 CC 133 -QAS-----YEP LG-----ISLRPLEAP--PTPTAPPGSP 161
 CC 121 GQAAGGQSLGPPASQNVNSHVKEPAGLEBEFVSRTLGIVPRDQPRGSHSPQRQLHSP 180
 CC 162 RRSSE-----HPDPTESR--SCSGSPSPSPDPKACNMWKYKITYV----- 202
 CC 181 AQSEGPSSLCGLKQALKRCPLEDEKKBPDG-KVPPRPLEAEGAOLQSGSNWEVVVOYED 239

RC STRAIN=CD-1; TISSUE=Testis;
 RX MEDLINE=96207307; PubMed=8617494;
 RA Shannon M., Ashworth L.K., Mucenksi M.L., lamerdin J.B., Branscomb E.,
 RT Scudde L.;
 CC "Comparative analysis of a conserved zinc finger gene cluster on human
 RT Chromosome 19q and mouse chromosome 7.";
 RL Genomics 33:112-120(1996).
 CC -|- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
 CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -|- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -|- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: U6186; AAB03529.1; -.
 DR HSSP: P25490; 12NM.
 DR MGD: MG1:107611; Zfp93.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 13.
 DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PRO0046; ZINCFINGER.
 DR ProDom: PD000003; Znf_C2H2; 11.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; Znf_C2H2; 13.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 13.
 DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 13.
 DR TRANSCRIPTION regulation; DNA-binding; Zinc-finger; Metal-binding;
 KW Nuclear protein; Repeat.
 FT DOMAIN 8 86 KRAB.
 FT ZN_FING 285 643 ZINC_FINGERS.
 FT ZN_FING 285 307 C2H2-TYPE.
 FT ZN_FING 313 335 C2H2-TYPE.
 FT ZN_FING 341 363 C2H2-TYPE.
 FT ZN_FING 369 391 C2H2-TYPE.
 FT ZN_FING 397 419 C2H2-TYPE.
 FT ZN_FING 425 447 C2H2-TYPE.
 FT ZN_FING 453 475 C2H2-TYPE.
 FT ZN_FING 481 503 C2H2-TYPE.
 FT ZN_FING 509 531 C2H2-TYPE.
 FT ZN_FING 537 559 C2H2-TYPE.
 FT ZN_FING 565 587 C2H2-TYPE.
 FT ZN_FING 593 615 C2H2-TYPE.
 FT ZN_FING 621 643 C2H2-TYPE.
 SO SEQUENCE 645 AA; 73012 MW; 7459CD140F5AF469 CRC64;

Query Match 15.3%; Score 427.5; DB 1; Length 645;
 Best Local Similarity 37.3%; Pred. No. 3e-15;
 Matches 91; Conservative 29; Mismatches 87; Indels 37; Gaps 6;

QY 266 FRCG---APASPYLLTSQAQDTSGSPSRARPLDSEFFSCQNEAVAGCGSGD--- 318
 DB 369 YKGGDCGKRFSSCSNHTHGRVHTEKP-----YECNECGKRFSLSGNLDIHQ 416
 QY 319 SLVPGDEDEPKYKOLCGSSFRYGNLASHRTVATGEKPYHCISGAFPNRPAVLKHSRI 378
 DB 417 RYVITG--EKRYKBECKGKGFSSASSQSHORVATGEKPFHCISGAKNFSHSHFLDHQRI 474
 QY 379 HSGEKPYKCECTCSR---VQVNSQPPSGFGKPARAGVGQKGFSSQRODLKSPPSQV 435
 DB 475 HTGEKPYRCEVCGRRFPMWSLSLHSHQSVHTGKPKYCGEGCKGF-----SHA 521
 QY 436 ARLRAVLLITGEKPYPCPPCGRRFRLLOTLSKSHVRIHTGEKPYHCDPCGLLRHKSQRL 495
 DB 522 SSLQAHSHSVATGEKPFKCNVCOKOFKSTSNLQAHQHVHTGEKPYKCDTGCKAFSSQKSLQ 581

QY 496 LHLR 499
 DB 582 VHQR 585

RESULT 9
 ZN85_HUMAN
 ID_ZN85_HUMAN STANDARD: PRT; 595 AA.
 AC 003923;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Zinc finger protein 85 (Zinc finger protein HPF4) (HTRF).
 GN ZNF85.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Placenta;
 RX MEDLINE=99053537; PubMed=9839802;
 RA Poncelet D.A., Bellefroid E.J., Bastiaens P.V., Demotie M.A.,
 RA Marine J.C., Pendeville H., Alami Y., Devos N., Lecocq P.J., Ogawa T.,
 RA Muller M., Martial J.A.;
 RT "Functional analysis of ZNF85 KRAB zinc finger protein, a member of
 RT the highly homologous ZNF91 family.";
 RL DNA Cell Biol. 17:931-943(1998).
 RN [2]
 RP SEQUENCE OF 1-196 FROM N.A.
 RX MEDLINE=91219421; PubMed=2023909;
 RA Bellefroid E.J., Poncelet D.A., Lecocq P.J., Revelant O.,
 RA Martial J.A.;
 RT "The evolutionarily conserved Kruppel-associated box domain defines a
 RT subfamily of eukaryotic multifingered proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 88:3608-3612(1991).
 CC -|- FUNCTION: TRANSCRIPTIONAL REPRESSOR.
 CC -|- SUBCELLULAR LOCATION: Nuclear.
 CC -|- TISSUE SPECIFICITY: EXPRESSED PREFERENTIALLY IN TESTICULAR
 CC TISSUES.
 CC -|- DEVELOPMENTAL STAGE: EXPRESSED EARLY DURING EMBRYONIC DEVELOPMENT.
 CC -|- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS.
 CC -|- SIMILARITY: CONTAINS 1 KRAB DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)
 CC -----
 CC EMBL: U35376; AAA79179.1; -.
 DR EMBL: M61866; AAA52689.1; -.
 DR EMBL: M61868; AAA58671.1; -.
 DR PIR: A39384; A39384.
 DR PIR: C39384; C39384.
 DR HSSP: P08048; 7ZNF.
 DR TRASPAC: T04990; -.
 DR Genew: HGNC:13160; ZNF85.
 DR MIM: 603899; -.
 DR InterPro: IPR001909; KRAB.
 DR InterPro: IPR000822; Znf_C2H2.
 DR Pfam: PF00096; zf-C2H2; 15.
 DR Pfam: PF01352; KRAB; 1.
 DR PRINTS: PRO0048; ZINCFINGER.
 DR ProDom: PD000003; Znf_C2H2; 13.
 DR SMART: SM00349; KRAB; 1.
 DR SMART: SM00355; Znf_C2H2; 15.
 DR PROSITE: PS00805; KRAB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 14.

DR PROSITE; P850157; ZINC_FINGER_C2H2_2; 15.
KW Transcription regulation; Zinc-finger; DNA-binding; Metal-binding;
Nuclear protein; Repeat; Repressor.
FT DOMAIN 4 75 KRAM.
FT ZN_FING 146 588 ZINC_FINGERS.
FT ZN_FING 146 168 C2H2-TYPE (DEGENERATE).
FT ZN_FING 174 156 C2H2-TYPE.
FT ZN_FING 202 224 C2H2-TYPE.
FT ZN_FING 230 252 C2H2-TYPE.
FT ZN_FING 258 280 C2H2-TYPE.
FT ZN_FING 266 308 C2H2-TYPE.
FT ZN_FING 314 336 C2H2-TYPE.
FT ZN_FING 342 364 C2H2-TYPE.
FT ZN_FING 370 392 C2H2-TYPE.
FT ZN_FING 398 420 C2H2-TYPE.
FT ZN_FING 426 448 C2H2-TYPE (DEGENERATE).
FT ZN_FING 454 476 C2H2-TYPE.
FT ZN_FING 482 504 C2H2-TYPE.
FT ZN_FING 510 532 C2H2-TYPE.
FT ZN_FING 538 560 C2H2-TYPE.
FT ZN_FING 566 588 C2H2-TYPE.
FT CONFLICT 84 84 R -> Q (IN REF. 2; AA552689).
FT CONFLICT 115 115 T -> I (IN REF. 2; AA58671).
FT CONFLICT 177 177 T -> R (IN REF. 2; AA58671).
FT CONFLICT 184 184 G -> R (IN REF. 2; AA58671).
SQ SEQUENCE 595 AA; 68718 MW; 44AA0A236D2D43B CRC64;
Query Match 15.2%; Score 424; DB 1; Length 595;
Best Local Similarity 38.5%; Pred. No. 4.1e-15;
Matches 90; Conservative 26; Mismatches 92; Indels 26; Gaps 5;
QY 267 KCGAPASTPYLLTSQADPTSGSPERARPLPGSEFPSCQNCVAVAGCSSGL---DSLVPG 323
DB 206 EGG-----KAFNMSSTLTTHKRIHTGKPKYKCEGAFNQSNNLTKHKIHTG 254
QY 324 DEDKPYCOLCRSSFFRYKGNLASHRTVHTGKPYHCSICGAPFNRPANLKTSHRIHSGK 383
DB 255 --EKPKYKCEGAFNRFSTLTTHKRIHTGKPKYKCEGAFNRSSTLTTHKIKITGK 312
QY 384 PYKCEGSAFVQVRSGPFGKRGVGVGKGFCSNQDLSPPSQVAHLPAHL 443
DB 313 PYKCEGSAFVQVRSGPFGKRGVGVGKGFCSNQDLSPPSQVAHLPAHL 443
QY 444 IHTGKPYCPCTGTRFRLIQLTKSHVRIHTGKPYHCDYHGLFPHKSQLRLH 497
DB 363 IHTGKPYKCEGAFNRFSTLTTHKRIHTGKPKYKCEGAFNRSSTLTTHK 362
RESULT 10
ID 2239 HUMAN STANDARD; PRT; 458 AA.
AC 016600;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Zinc finger protein 239 (Zinc finger protein MOK-2) (HOK-2).
GN ZNP239 OR MOK2 OR HOK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=96139011; PubMed=8587123;
RT "Human and mouse Kruppel-like (MOK2) orthologue genes encode two
RT different zinc finger proteins."
RL J. Mol. Evol. 41:784-794(1995).
CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential)
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X82126; CAA57638.1; -
DR EMBL; X82125; CAA57637.1; -
DR HSSP; P25490; IUBD.
DR Genew; HGNC:13031; ZNF239.
DR MIM; 601069; -
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00096; zf-C2H2; 5.
DR PRINTS; PR00048; ZINC_FINGER.
DR PRODOM; PD000003; Znf_C2H2; 7.
DR SMART; SM00355; Znf_C2H2; 9.
DR PROSITE; P850028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; P850157; ZINC_FINGER_C2H2_2; 9.
KW Transcription regulation; DNA-binding; Zinc-finger; Metal-binding;
Nuclear protein; Repeat.
FT DOMAIN 235 453 ZINC_FINGERS.
FT ZN_FING 235 257 C2H2-TYPE.
FT ZN_FING 263 285 C2H2-TYPE.
FT ZN_FING 291 313 C2H2-TYPE.
FT ZN_FING 319 341 C2H2-TYPE.
FT ZN_FING 347 369 C2H2-TYPE.
FT ZN_FING 375 397 C2H2-TYPE.
FT ZN_FING 403 425 C2H2-TYPE.
FT ZN_FING 431 453 C2H2-TYPE.
SQ SEQUENCE 458 AA; 51534 MW; 7749A0629A9D8920 CRC64;
Query Match 15.1%; Score 421.5; DB 1; Length 458;
Best Local Similarity 30.5%; Pred. No. 4.3e-15;
Matches 107; Conservative 32; Mismatches 129; Indels 83; Gaps 10;
QY 176 SCGQPPSPASDPACN-----WKRYKYLINSQA-----SQAGLVGRSS 218
DB 137 TQNG-QLKESLDIPDCKCKIHWK--SQVSCGQGHTEKPCDINNGKILNTSPD 193
QY 219 GQPCQARLPSCDEASSSSSSSSSSEGPISGFSLSPTAATVQKCAPASTPYLL 278
DB 194 GHPYKIKITAEKQVGGSCGKNFSQSS- 223
QY 279 TQADPTGSPERARPLPGSEFPSCQNC-EAVAGCSSGLDVLVGDDEPKYCOLCRSS 337
DB 224 LHQDHTKRP-----YKCEGCGKFTSSSLIHQAVHTDEKPKYKCEGCGK 271
QY 338 FRYKGNLASHRTVHTGKPYHCSICGAPFNRPANLKTSHRIHSGKPYKCEGSAFVQ- 396
DB 272 FTRSSSLIHAVHTGKPYKCDKCGKPSQSKLHQRVHTGKPYKCEGSAFVQ 331
QY 397 ---VRSQPPSGFGKPARGVGVGKGFCSNQDLSPPSQVAHLPAHLIHTGKPY 452
DB 332 SWLHIHQDHTG--ERPKYKCEGCGK- 427
QY 453 CPCTGTRFRLIQLTKSHVRIHTGKPYHCDYHGLFPHKSQLRLH 503
DB 377 CYECGKPSQSDLRHLHVRHTGKPYHCGKGFSSSKLILHQRVHTG 427
RESULT 11
ID 2145 HUMAN STANDARD; PRT; 673 AA.
AC Q05516;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Zinc finger protein PLZF (Promyelocytic leukemia zinc finger
DE protein) (Zinc finger protein 145).
GN ZNF145 OR PLZF.
OS Homo sapiens (Human).

OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
OX	NCBI_TaxId=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Heart ventricle;	
RX	MEDLINE=93209216; PubMed=834553;	
RA	Chen S.-J., Zelen A., Tong J.-H., Yu H.-Q., Wang Z.-Y., Derre J.,	
RA	Berger R., Waxman S., Chen Z.;	
RT	"Rearrangements of the retinoic acid receptor alpha and promyelocytic	
RT	leukemia zinc finger genes resulting from t(11;17) (q23;q21) in a	
RT	patient with acute promyelocytic leukemia.";	
RL	J. Clin. Invest. 91:2260-2267(1993).	
RN	[3]	
RP	X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 7-122.	
RX	MEDLINE=20005701; PubMed=10537309;	
RA	Li X., Peng H., Schultz D.C., Lopez-Guiza J.M., Rauecher F.J. III,	
RA	Marmorstein R.;	
RT	"Structure-function studies of the BTB/POZ transcriptional repression	
RT	domain from the promyelocytic leukemia zinc finger oncoprotein.";	
RL	Cancer Res. 59:5275-5283(1999).	
CC	-1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. MAY PLAY A ROLE IN	
CC	MELOID MATURATION AND IN THE DEVELOPMENT AND/OR MAINTENANCE OF	
CC	OTHER DIFFERENTIATED TISSUES.	
CC	-1- SUBCELLULAR LOCATION: Nuclear.	
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PLZF2 AND PLZF2 (SHOWN HERE);	
CC	ARE PRODUCED BY ALTERNATIVE SPLICING.	
CC	-1- TISSUE SPECIFICITY: WITHIN THE HEMATOPOIETIC SYSTEM, PLZF2 IS	
CC	EXPRESSED IN BONE MARROW, EARLY MELOID CELL LINES AND PERIPHERAL	
CC	BLOOD MONONUCLEAR CELLS. ALSO EXPRESSED IN THE OVARY, AND AT LOWER	
CC	LEVELS, IN THE KIDNEY AND LUNG.	
CC	-1- INDUCTION: BY RETINOIC ACID.	
CC	-1- DISEASE: A FORM OF ACUTE PROMYLOCYTIC LEUKEMIA (APL) IS	
CC	CHARACTERIZED BY A CHROMOSOMAL TRANSLOCATION T(11;17) (Q32;Q21)	
CC	WHICH INVOLVES ZNF145 AND RETINOIC ACID RECEPTOR ALPHA (RAR).	
CC	-1- SIMILARITY: BELONGS TO THE KREPEL-FAMILY OF C2H2-TYPE ZINC-	
CC	FINGER PROTEINS.	
CC	-1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/)	
CC	or send an email to license@ib-sib.ch)	
DR	EMBL; Z19002; CA79489.1; -	
DR	EMBL; M60093; AAC60590.2; -	
DR	PDB; 1CS3; 09-AUG-00.	
DR	TRANSFAC; T02336; -	
DR	GENEW; HGNC:12930; ZNF145.	
DR	MIM; 176797; -	
DR	InterPro; IPR000210; BTB_POZ.	
DR	InterPro; IPR000822; Znf_C2H2.	
DR	Pfam; PF00096; Zf-C2H2; 9.	
DR	Pfam; PF00651; BTB; 1.	
DR	PRINTS; PR00048; ZINC_FINGER.	
DR	ProDom; PD000003; Znf_C2H2_1.	
DR	SMART; SM00225; BTB; 1.	
DR	SMART; SM00355; Znf_C2H2; 9.	
DR	PROSITE; PS50097; BTB; 1.	
DR	PROSITE; PS00026; ZINC_FINGER_C2H2_1; 8.	
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.	

[illegible]

RESULT 12
 ID HIC1 HUMAN STANDARD; PRT; 733 AA.
 AC 014536;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 GN Hypermethylated in cancer 1 protein (Hic-1).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 [1]
 RP MEDLINE=96071505; PubMed=7585125;
 RA Waleis M.M., Biel M.A., el Dely W., Nelkin B.D., Isea J.-P.,
 RA Cavenne W.K., Kuerbitz S.J., Baylin S.B.,
 RT "p53 activates expression of Hic-1, a new candidate tumour suppressor
 gene on 17p13.3."
 RL Nat. Med. 1:570-577(1995).
 [2]
 RP ALTERNATIVE SPLICING, INTERACTION WITH CTBP, AND SUBCELLULAR LOCATION.
 RX MEDLINE=21438979; PubMed=11554746;
 RA Delcourt S., Plante S., Guerdard C., Leprieux D.,
 RT "Characterization of HRC22, a human homologue of the putative tumor
 suppressor gene HIC1."
 RL Biochem. Biophys. Res. Commun. 287:427-434(2001).
 CC -1- FUNCTION: Transcriptional repressor. May act as a tumor
 suppressor. May be involved in development of head, face, limbs
 and ventral body wall.
 CC -1- SUBUNIT: Interacts with CTBP.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- ALTERNATIVE PRODUCTS: At least 2 isoforms: 1 (shown here) and 2;
 are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed with highest levels
 found in lung, colon, prostate, thymus, testis and ovary.
 CC Expression is absent or decreased in many tumor cells.
 CC -1- DISEASE: Defects in HIC1 may be a cause of the facial dysmorphism
 and additional birth defects (except for lisencephaly) seen in
 the contiguous gene disorder Miller-Dieker syndrome (MDS), like
 defective development of the nose, jaws, extremities,
 and gastrointestinal tract, and kidney.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
 FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 C2H2-TYPE ZINC FINGERS.
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 DR EMBL: I41919; AAD09201.1; -
 DR HSSP: P08046; 1A1H.
 DR Genew: HGNC:4909; HIC1.
 DR MIM: 603825; -
 DR InterPro: IPR000210; BTB_POZ.
 DR InterPro: IPR000822; ZnF_C2H2.
 DR Pfam: PF00651; BTB_1.
 DR Pfam: PF00956; ZF-C2H2; 5.
 DR PRINTS: PR00048; ZINC_FINGER.
 DR PRODOM: PD000003; ZnF-C2H2; 1.
 DR SMART: SM00325; BTB; 1.
 DR SMART: SM00355; ZNF_C2H2; 5.
 DR PROSITE: PS50097; BTB; 1.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 5.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 5.

KW Multigene family; Nuclear protein; Transcription regulation;
 KW DNA-binding; Zinc-finger; Metal-binding; Repeat; Anti-oncogene;
 KW Developmental protein; Alternative splicing.
 FT DOMAIN 47 110
 FT DOMAIN 160 167
 FT DOMAIN 195 199
 FT DOMAIN 439 613
 FT ZN_FING 439 459
 FT ZN_FING 509 529
 FT ZN_FING 537 557
 FT ZN_FING 565 585
 FT ZN_FING 593 613
 FT SITE 241 245
 FT VARSPLIC 1 19
 SQ SEQUENCE 733 AA; 76467 MW; E8034428A4E490D3 CRC64;
 MISSING (IN ISOFORM 2).
 Query Match 15.1%; Score 420.5; DB 1; Length 733;
 Best Local Similarity 23.9%; Pred. No. 7,6e-15;
 Matches 151; Conservative 54; Mismatches 201; Indels 227; Gaps 20;
 QY 20 HSSDVLGNLNLRLRGLTDLVTLVGGQPLRAHKAVLJACSGFPYSIFRGAGVGVDS 79
 DB HSRQLQLQNLNORTGFLCDVITVQNALFRHKVLAASAYLKSIV-----VADNLN 83
 QY 80 LFGGPEANG-FAPLIDFMYTSL-----RLSPATAP--AVLAATYLAQMEHVQ 125
 DB LDHMDVSPAVFRLVDFIYTGRLADGAEAAAVAFAPGAPBLGAVLAAASVLOIPDLVA 143
 QY 126 ACHRFIQ-----ASYEPLGISIRPLEAR-----PYTPPTAP 157
 DB LCKKRLKHKGYCHLRGGGGGGGYAPYGRPRGRLAATPVYQACYSRSPVGPAPPAEP 203
 QY 144 LCKKRLKHKGYCHLRGGGGGGGYAPYGRPRGRLAATPVYQACYSRSPVGPAPPAEP 203
 DB PGSPRS-----EGHPDP-----TESRCS-----OGPPSPAPSD 188
 QY 204 PGSPRAAVNTHCALYASGPAPPAALCASERCSPGLGDSKSPPGSAAPERPLABRE 263
 DB 189 -----PRACNMWKYK-----YVLNSAQAGSLVGRSSGQPCQARL 227
 QY 264 LPPRDSPPSAGAPAYKPPPLALPPLPPLPQKLEAPAPSPFRFGSGSPDEPRGRPD 323
 DB 228 -PS-----GDEASSSSSSSSSSSEEG-----PIPGQSRLSPYATVQF 266
 QY 324 GSLLYRWKIEPPGLSGYDELRGRGSPSECEBGRGDAAVSPGPGPLGLAPPPRYGS 383
 DB 267 KCGAFA---STPYLTLSQAQDTSSGSPS---RARPPLP-----GSFFECQNC- 307
 QY 384 LDGPGAGDGDYKSSSBETGSSSEDPSPGGLLEGYPCHLAYGPSP3FGDNLVYCTPCG 443
 DB 308 -----BAVACSSGLDSL 320
 QY 444 KQFPSSSEQLNAVEAAVEEENLYGARAABAAGLPGPPGGG3DKVAGAPGLGL 503
 DB 321 VEGDEDKPKYKQLCSSFRFYKGNLASHTVTGKPYHCISICGAPFNPANLKTSHRS 380
 QY 504 L-----RPYRCASCDSKYKDPATLRQHEKTHMLTRPYCTICGKFFRTGTRHMRSL 558
 DB 381 GEKPYKCEGSRFRQVNSQPPSGRGKPRAGCVGQKGFCSGQKDLSPSSQVAHLRA 440
 QY 559 GLKPPACACGMR-----TROYRLTE 580
 DB 441 HVLITGEKPYPCPTGTRPRLIOTLKSHVRH 473
 DB 581 HMRHSGEKPYECQYCGGKFAQGRNLISHMKH 613
 -----TROYRLTE 580
 RESULT 13
 ID HIC2 HUMAN STANDARD; PRT; 615 AA.
 AC 096783; Q9UPX3; Q9NSM9; Q9EKR3;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Hypomethylated in cancer 2 protein (Hic-2) (Hic-3) (Hic1-related gene
on chromosome 22).
GN HIC2 OR HIC22 OR KIAA1020.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=21438975; PubMed=11554746;
RX MEDLINE=21438975; PubMed=11554746;
RA Deltour S., Pitte S., Guerardel C., Lepince D.;
RT "Characterization of HIC22, a human homologue of the putative tumor
suppressor gene Hic1.";
RL Biochem. Biophys. Res. Commun. 287:427-434(2001).
RN [2]
RP TISSUE=Brain;
RC "Complete deduced structure of Hic-3, a novel human bcl/poz and Zp
factor of the Hic family.";
RT Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=2057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA Camp M., Smink L.J., Alencough R., Almeida J.P., Babbage A.K.,
RA Bagshaw C., Bailey J., Barlow K.F., Bates K.N., Beasley O.P.,
RA Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA Burdill M.D., Burton J., Carter C., Carter N.P., Chen Y., Clark G.,
RA Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA Hall R.E., Hall-Tamlyn G., Heathcote R.W., Ho S., Holmes S.,
RA Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA Laird G.K., Langford C.F., Levertha M.A., Lloyd C., Lloyd D.M.,
RA Maryn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA McClellan J., McLaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA Phillips S.H., Plumb R.W., Ramsey H., Ramsey Y., Rogers L., Ross M.T.,
RA Scott C.B., Sena H.K., Stuce C.D., Smalley S., Smith W.L.,
RA Staudt R.C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA Vaudin M., Wall M., Wallis J.W., Whiteley M.N., Willey D.L.,
RA Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kodoh J.,
RA Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuizawa S.,
RA Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S., Do A., Do T.,
RA Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S., Sloan D., Song L.,
RA Wang Q., Wang Y., Wang Z., White J., Williamson D., Wu H., Yao Z.,
RA Zhan M., Zhang G., Chiswick S., Murray J., Miller N., Minx P.,
RA Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA Hinde K., Kemp K., Latreille P., Layman D., Ozerick P., Rohlfing T.,
RA Korfi I., Bebell J.A., Hillier L., Mardis E., Waterston R., Wilson R.,
RA Emanuel B.S., Shaikh T., Kurahashi H., Salita S., Budarf M.L.,
RA McDevitt H.B., Johnson A., Wong A.C.C., Morrow B.E., Edelmann L.,
RA Kim U.J., Shizuya H., Simon M.I., Dumanski J.P., Peyrard M., Kedra D.,
RA Serousi E., Franssen I., Tapia I., Brudner C.E., O'Brien K.P.,
RA Wilkerson P., Bodenteich A., Hartman K., Hu X., Khan A.S., Lane L.,
RA Tikhunov Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
RN [4]
RP SEQUENCE OF 61-615 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirosewa M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;

RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:197-205(1999).
RN [5]
RP SEQUENCE OF 301-615 FROM N.A.
RC TISSUE=Testis;
RA Koehner K., Beyer A., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Transcriptional repressor.
CC -1- SUBUNIT: Interacts with CtBP.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 2 isoforms, 1 (shown here) and 2; are
CC produced by alternative splicing.
CC -1- TISSUE SPECIFICITY: Highest levels in cerebellum.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ13204; CAC70715.1; -;
DR EMBL: AF348035; AAK72951.1; -;
DR EMBL: AF000557; -; NOT_ANNOTATED_CDS.
DR EMBL: AB028943; BAA82972.1; -;
DR EMBL: AL162003; CAB82344.1; -;
DR Genew; HGNC:18595; HIC2.
DR Interpro: IPR000210; BTB_POZ.
DR Interpro: IPR000822; Znf_C2H2.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF00096; ZF_C2H2; 5.
DR SMART: SM00355; ZNF_C2H2; 5.
DR PROSITE: PS50097; BTB; 1.
DR PROSITE: PS50028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 5.
KW Multigene family; Nuclear protein; Transcription regulation;
KW DNA-binding; Zinc-finger; Metal-binding; Repeat; Repressor;
KW Alternative splicing.
FT DOMAIN 46 109 BTB.
FT ZN_DOMAIN 442 615 ZINC_FINGERS.
FT ZN_FING 442 469 C2H2-TYPE.
FT ZN_FING 505 532 C2H2-TYPE.
FT ZN_FING 533 560 C2H2-TYPE.
FT ZN_FING 561 588 C2H2-TYPE.
FT ZN_FING 589 615 C2H2-TYPE.
FT SITE 246 250 BINDING TO CTBP.
FT VAR_SEQ 1 18 MISSING (IN ISOFORM 2).
FT CONFLICT 49 51 ITM -> TIR (IN REF. 1).
FT CONFLICT 176 176 Q -> R (IN REF. 4).
SQ SEQUENCE 615 AA; 66155 MW; B03686C31B198C95 CRC64;
Query Match 15.0%; Score 419; DB 1; Length 615;
Best Local Similarity 22.9%; Pred. No. 7,7e-15;
Matches 153; Conservative 64; Mismatches 205; Indels 246; Gaps 20;
QY 6 APEGALGVREPTRSSDVLGNLNEIRLRLGILTDVTLVGGQPLRAKAVLIACSGFFYS 65
DB 14 AARGDGMPEMLPSHSKQLQLQNGRKGLCDVIIVWBNIPFAHGVLAASITPFS 73
QY 66 --IFRGAGVGVDTLSLPGEPARGFAPLDPMYTSRLSPATPA-----VLAATY 117
DB 74 LVLDHNLINLIDTDVMS-----STVFQOILDFITYG--KLPSDPAEPNFTLLTAASY 125
QY 118 LGMHEVYQAQCHRFQASVEPIGISLRPLEAEPPTPTAPPSPR-----RSEGHP 168
DB 126 LQLEPULALCRKRLKAGKPFQ-----SGRAGSTGMGRPPRSQRLSTRASVIAQAYQGLV 179


```

QY 169 DPTESRSCSGPSPSPAPDPKACMKKYYIVLNSQASQSLVGERSSGCPPOARIP 228
DB 180 DSRKHAHAFQELPOAKSDD-----ELFLGS---NDSVQGLGRAVCP 220
QY 229 SGGDEASSSSSSSSSSSEEG-----PI-----PGQSLSTFATVQPKCA 270
DB 221 AGGEGAGLGGSSSTNGSGCGEOELDLKSKSPPLPAPTPGHLLTPDDAQLSDSOHS 280
QY 271 P--ASTPYLLTS-----280
DB 281 PPAASAPYANASAYSELGTPDEPMDLGAEDNHLLEAGCGOPRKSLSHSTRKMG 340
QY 281 QADOTSGSEPERAR-----PLGSEF-----PSCNCEA 309
DB 341 KKEPVAAGSPFEREARPKPGCEGEGVGDVPNGILASGAGSPGYEPPEPKCE--BE 399
QY 310 VAGCSGGLDLYVPGD-----DKPYKQLCGSSFRYKCNLAS 346
DB 400 ENGKDASEDSAGSGSGSGSHASAHYMYROGYETVSYDNLVYCIPCAAGPSSBQLNA 459
QY 347 HRTVHTGE-----KPYHCSICGARFNRPAN 371
DB 460 HVEHTHEELFTKEGAVETGSGGABEADLSAPSAATTAERPRKCVCEKTYDPA 519
QY 372 LKTHSRHSGEKPYKCTCGSRFVYRSQPPSGFGKPARGVGKGFCSSQRODLKSP 431
DB 520 LKQHEKHWLTPFPNCIGKMETQ-----RG-----546
QY 432 PQGVATLRAHVLIHGEKPYPCPTCTRRHLOTLKSHVRITGKPYHCDPGLHFRK 491
DB 547 -TWTRHMRSHL---GLKPFACDECMRPTROYLRTEHMRVHSGEKPYECOLCGKFTQ 601
QY 492 SOLRLHLR 499
DB 602 RNLISHLR 609

RESULT 14
HIC1 CHICK STANDARD; PRT; 676 AA.
ID HIC1 CHICK STANDARD; PRT; 676 AA.
AC Q90850; Q90851; Q90852;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypermethylated in cancer 1 protein (Hic-1) (GammaFBP) (Fragment).
GN HIC1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OC NCBI_Taxid=9031;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC TISSUE=Embryonic lens;
RA MEDLINE=94374565; PubMed=8088434;
RX Liu Q., Shalaby F., Puri M.C., Tang S., Breitman M.L.;
RT "Novel zinc finger proteins that interact with the mouse gamma
RT F-cysteinein promoter and are expressed in the sclerotic during early
RT somatogenesis.";
RL Dev. Biol. 165:165-177(1994).
CC -1- FUNCTION: Binds specifically to the gamma F-1-binding motif of the
CC gamma F-cysteinein promoter. May have a regulatory role in
CC sclerotic specification and/or differentiation. Isoform 2
CC functions as a transcriptional repressor in lens cells.
CC -1- SUBUNIT: Interacts with CtBP (by similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS: At least 3 isoforms; 1/GammaFBP-A,
CC 2/GammaFBP-B and 3/GammaFBP-C (shown here); are produced by
CC alternative splicing.
CC -1- TISSUE SPECIFICITY: Isoform 1 is highly expressed in kidney and
CC lung. Expression of isoform 2 is higher in the lens, retina and
CC stomach, and extremely low in heart, muscle, kidney and lung.
CC Isoform 3 is weakly expressed in heart, kidney and lens.

```

```

CC -1- DEVELOPMENTAL STAGE: In the embryo of stage 11, expressed
CC predominantly in the head mesenchyme surrounding the brain and in
CC the paraxial mesoderm. Highly expressed in presomitic mesoderm and
CC then over the entire epithelial somite. During somitic
CC differentiation, expression becomes restricted to the sclerotic.
CC In the developing lens, expression is most active at the beginning
CC of lens fiber cell differentiation.
CC -1- SIMILARITY: BELONGS TO THE KRUEPPEL FAMILY OF C2H2-TYPE ZINC-
CC FINGER PROTEINS. HIC TRANSCRIPTION FACTORS SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL, X79011; CAAS5644.1; -
DR EMBL, X79050; CAAS5652.1; -
DR EMBL, X79051; CAAS5653.1; -
DR HSSP, P08046; 1A1H.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF00096; zfc_C2H2; 5.
DR PRINTS; PR00048; ZINCFINGER.
DR PRODOM; PD000003; Znf_C2H2; 1.
DR SMART; SM00225; BTB; 1.
DR PROSITE; PS00097; BTB; 1.
DR PROSITE; PS00355; Znf_C2H2; 5.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 5.
DR Nucleic acid protein; Transcription regulation; DNA-binding; Zinc-finger;
DR Metal-binding; Repeat; Developmental protein; Repressor;
KW Alternative splicing.
FT NON TER 1
FT DOMAIN 1 63 126 BTB.
FT DOMAIN 420 585 ZINC FINGERS.
FT ZN FING 420 585 C2H2-TYPE.
FT ZN FING 474 501 C2H2-TYPE.
FT ZN FING 502 529 C2H2-TYPE.
FT ZN FING 530 557 C2H2-TYPE.
FT ZN FING 558 585 C2H2-TYPE.
FT SITE 241 245 BINDING TO CTBP.
FT VARSPLIC 1 35 APGARPARSRGRGKRSREERCGEAGAGRRARGA -> MR
FT VARSPLIC 1 35 VHRLEGLWAEBSGRGRARGA (IN ISOFORM 2).
FT SEQUENCE 676 AA; 73758 MW; 3D3C7PD3302F32EC CRC64;

Query Match 15.0%; Score 419; DB 1; Length 676;
Best Local Similarity 23.7%; Pred. No. 8.4e-15;
Matches 146; Conservative 67; Mismatches 184; Indels 218; Gaps 21;

QY 9 GAGVGVRFTSSSVLGNLNLRLRGILTVTLVGGQPLRAHRAVLIAGCFYSIFR 68
DB 34 GMLFAMEVPSHROLTLQTLTQRTKGLCDYIVQNALFRANNTILAAASAVKSLV- 92
QY 69 GAGVGVDLSLPGGPEARG-FAPLLDPMYISRL-RLSPA---TAPAVLIAATYLOMEHV 123
DB 93 ---VHDVLNLDHMSVPGIFRLIDFTYGRGCEPGSGBSGAVLAASVYQIDGL 148
QY 124 VOACHR-----PIQASYE--PLGISLRPLR-----146
DB 149 VALCKKLRSGKGYCHLRGAYAYKLGRLATTPVIOACVSGTFRPVLDLPVEPAAPLN 208
QY 147 -----A 147
DB 209 TOGGLVASAGCPPLPHGLCPERRHCSPPCGLDLSKSPGSAJLLPTDRLLPAEPR 268
QY 148 EPTFP-----TAPPGSPRRSGH---DPTESRSCSGPSP 184

```

Db. 269 EPELPRHSDPPVSGGLLGHAPAYKDSPEGCEP---GCHPMTDPRSTPPCAE-PPPE 324
 QY 185 ASDPDKAC-NMKYK----YIVNSQASQAGSLVGRSSQCP--CPOARLPs--GDEASS 235
 Db 325 RGGRELMYRMHHEPILGPYL---DEGEAKELEREKASPPAAQOPRIPSPESNDLEP 381
 QY 236 SSSSSSSSSSECEPIPG-----POGRSPPTATVQFKCGAPASTPYLLTSGA 282
 Db 382 DNSTSERTSGSEGPSPDALDRKCNHLYGPESELGDLVYCIPCGKGFPESEGLNHAVEA 441
 QY 283 QDTS---GSPSRARLPSESEFFSCONCAVAGCSSGLDVLPGDEDKYKQOLCRSSP 338
 Db 442 HNEEELHYKKAADAVP-----LDKGAGL-----GILRPYRCSSCDKSY 493
 QY 339 RYKGNLASHRTVTGKPYHCSICGARFNPRLKTHSRHSGEKPYKCEGSRFYQVR 398
 Db 484 KDPATLRQHEKTMULRPYCTICGKKFTQGRMTNMRHLGKPPACDAGCMRP----- 539
 QY 399 SQPPSGFGKPARGVGQKGGFSSORQDKSPPOVAHLRAHVLHTEGKPYPCPTCGT 458
 Db 540 -----TQYRLTEHRIHSGEKPYECQVCG 565
 QY 459 RFRHLOTLSHVRIH 473
 Db 566 KFAQORNLISHMKKH 580

RESULT 15
 HIC1_MOUSE STANDARD; PRT; 892 AA.
 AC 09R1Y5; 09R1Y6; 09R2B0;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypermethylated in cancer 1 protein (Hic-1).
 GN HIC1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND DEVELOPMENTAL STAGE.
 RC STRAIN=129/SV, and Swiss Webster; TISSUE=Embryo;
 RX MEDLINE=99172081; Pubmed=10072440;
 RA Grimm C., Spoerle R., Schmid T.E., Adler I.-D., Adamski J.,
 RA Schughart K., Graw J.;
 RT "Isolation and embryonic expression of the novel mouse gene Hic1, the
 RT homologue of Hic1, a candidate gene for the Miller-Dieker syndrome";
 RL Hum. Mol. Genet. 8:697-710(1999).
 RN [2]
 RP SEQUENCE FROM N.A., DISEASE, AND TISSUE SPECIFICITY.
 RX MEDLINE=20122251; Pubmed=10655551;
 RA Carter M.G., Johns M.A., Zeng X., Zhou L., Zink M.C., Mankowski J.L.,
 RA Donovan D.M., Baylin S.B.;
 RT "Mice deficient in the candidate tumor suppressor gene Hic1 exhibit
 RT developmental defects of structures affected in the Miller-Dieker
 RT syndrome";
 RL Hum. Mol. Genet. 9:413-419(2000).
 RN [3]
 RP SEQUENCE OF 179-338 FROM N.A.
 RC STRAIN=129/SV, TISSUE=Liver;
 RX MEDLINE=99297610; Pubmed=10371200;
 RA Guerdel C., Delcours S., Leprieux D.;
 RT "Evolutionary divergence in the broad complex, tramtrack and bric a
 RT brac/poxviruses and zinc finger domain from the candidate tumor
 RT suppressor gene hypermethylated in cancer";
 RL FEBS Lett. 451:253-256(1999).
 CC -1- FUNCTION: Putative transcription factor. May act as a tumor
 CC suppressor. May be involved in development of head, face, limbs
 CC and ventral body wall.
 CC -1- SUBUNIT: Interacts with CtBP (by similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS: A number of isoforms may be produced by

CC alternative splicing.
 CC -1- TISSUE SPECIFICITY: Ubiquitously expressed with highest levels in
 CC heart and lung.
 CC -1- DEVELOPMENTAL STAGE: Expression is first detected in the embryo
 CC after 9 days post coitum. In the embryo, expression is found in
 CC restricted regions of somite derivatives, limb anlagen and cranio-
 CC facial mesenchyme. In the fetus, it is additionally expressed in
 CC mesenchymes apposed to precartilaginous condensations, at many
 CC interfaces to budding epithelia of inner organs, and weakly in
 CC muscles.
 CC -1- DISEASE: Defects in HIC1 are the cause of perinatal death with
 CC serious developmental anomalies, including acrania, exencephaly,
 CC cleft palate, omphalocele, craniofacial and limb anomalies.
 CC -1- SIMILARITY: BELONGS TO THE KRUEPEL FAMILY OF C2H2-TYPE ZINC-
 CC FINGER PROTEINS. HIC TRANSSCRIPTION FACTORS SUBFAMILY.
 CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
 CC -1- SIMILARITY: CONTAINS 5 C2H2-TYPE ZINC FINGERS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AF036334; AAD30654.1; -;
 CC DR EMBL; AF036582; AAD30655.1; -;
 CC DR EMBL; AF132691; CAB44493.1; -;
 CC DR HSSP; P08046; IAH1.
 CC DR MGD; MG1:138010; Hic1.
 CC DR InterPro; IPR000210; BTB_POZ.
 CC DR InterPro; IPR000822; Znf_C2H2.
 CC DR Pfam; PF00651; BTB; 1.
 CC DR Pfam; PF00096; Zf-C2H2; 5.
 CC DR PfDom; PD000003; Znf-C2H2; 1.
 CC DR SMART; SM00225; BTB; 1.
 CC DR SMART; SM00355; Znf_C2H2; 5.
 CC DR PROSITE; PS00097; BTB; 1.
 CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 5.
 CC DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 5.
 CC KW Multigene family; Nuclear protein; Transcription regulation;
 CC KW DNA-binding; Zinc-finger; Metal-binding; Repeat;
 CC KW Developmental protein; Alternative splicing.
 CC FT DOMAIN 1 172 ARG/PRO/GLY-RICH.
 CC FT DOMAIN 206 269
 CC FT DOMAIN 271 278 POLY-ALA.
 CC FT DOMAIN 319 326 POLY-GLY.
 CC FT DOMAIN 354 358 POLY-PRO.
 CC FT DOMAIN 596 777 ZINC FINGERS.
 CC FT ZN_FING 596 623 C2H2-TYPE.
 CC FT ZN_FING 666 693 C2H2-TYPE.
 CC FT ZN_FING 694 721 C2H2-TYPE.
 CC FT ZN_FING 722 749 C2H2-TYPE.
 CC FT ZN_FING 750 777 C2H2-TYPE.
 CC FT SITE 400 404 BINDING TO CTBP.
 CC FT SITE 242 242 N -> S (IN REF. 1; AAD30654).
 CC FT CONFLICT 242 242 N -> S (IN REF. 1; AAD30654).
 CC SQ SEQUENCE 892 AA; 94320 MW; BBD16AA051995740 CRC64;

Query Match 15.0%; Score 418; DB 1; Length 892;
 Best Local Similarity 23.2%; Pred. No. 1.2e-14;
 Matches 156; Conservative 53; Mismatches 198; Indels 266; Gaps 23;

QY 2 GSPAPREGALG-----YREFRRSSDVLGNINELRLGILLTDVLLVGGGELRAHKVYL 56
 Db 165 GLPTCPBEGCAGQMLDTEAPGHSRQLLDQNNORTGFLCDVITIVONALFRHKVYL 224
 QY 57 IACSGFFYSIFRGAGVGVVLSLPGCEPARG-FAPLDLPMYTSRL-----RLSP 105
 Db 225 AASSAYLKSIV-----VADNLNLNDHDMVSPAVPLVDFITGTSLTSVEAAAAAVAP 279
 QY 106 ATAP---AVLAATYLOMEHVAVOACHRFQ-----ASYEPGISLRPLEA 147

```
Db      280 GAPELGLAVLAASYLOIPDLVALCKKLIKHKHCKTCHLRGGSGGGGAYAPYGRGLRA 339
Qy      148 E-----PPTPTAPPGPSP-----161
Db      340 ATPVIQACSSPAGPPPPAAEPGSPDAVNTHCALYASGPAAACAPERRCSPLC 399
Qy      162 -----RRSEG 166
Db      400 GLDLCKSPGSSVPERPLSERELPPRPDSPPGAGPAVYKEPSLALPLPLPQKLEBA 459
Qy      167 HPDP-PTESRSCSQPPSPAPDPKAC--NMKKYKIVLNSQASQAGSLYGERSS-----218
Db      460 VFTDPFRSGSGSPGEPGPRPDGSSLLYRMKHR---?GLGSYGDELVRDRGSGPGERL 515
Qy      219 ---GQPC-----PQARLP-----6GDEASSSSSSSSSEEGPIPGQSR 256
Db      516 BERGGDPAASPGGPPGLVPPPRYPSLDGPGTGADGDDTKSSSEETGSSDDPSP-PGQH 574
Qy      257 LSPTAATVQFKCGAPASTPYLLTSAQD-----TSGSPS-----ERA 293
Db      575 LE-----GYPC--PHLAYGEPESEFGDNLVYCIPCGKGFPSEBQLNAHVEAHVEEE 622
Qy      294 RPLPGSEFFSCNCAVAVGCCSSGLDSLVPGDEDK-----PYKQQLCRSSFRY 340
Db      623 EALYg---RAEAAEVAAG-AAGLGPPFGGGGDKVTGAPGGLGELLRPYRCASCDKSYXD 677
Qy      341 KGNLASHRTVHTGEKPYHCSICGARFNRPAHLKTHSRIHSGEKPYPKCTCGSRFYQVRSQ 400
Db      678 PATLAQHEKTHWLTTPYCTICGKFTQKGTMTRHMRSHLGLKFPACDACGKRF-----731
Qy      401 PPSGFQGKPARGGVQKGGFCSSORODLKSPSQVAHLRLAHVLIHTGEKYPCTCGTRF 460
Db      732 -----TRQYRLTEHMRHISGEKPYECQVCGGKF 759
Qy      461 RHLQTLKSHVRTH 473
Db      760 AQQRNLISHMKH 772
```

Search completed: July 14, 2003, 18:16:07
Job time : 8.38782 secs

THIS PAGE BLANK (USPTO)

XX Novel angiogenesis associated polypeptides and polynucleotides encoding
 PT the polypeptides, useful for modulating angiogenesis and for treating
 PT tumors and cancers -
 XX
 PS Claim 1, Page 12-13, 159pp; English.
 XX
 CC The invention relates to angiogenesis associated proteins (AAP) and their
 CC corresponding cDNA molecules, which are useful for modulating
 CC angiogenesis. AAP proteins and nucleic acids are useful for promoting
 CC wound healing, for example after organ transplantation, and in the
 CC treatment of tumours, myocardial infarction, cancers, diabetic
 CC retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.
 CC AAP proteins and DNA's are useful in potential prophylactic and
 CC therapeutic applications implicated in a variety of disorders including
 CC those related to angiogenesis, and also in diagnostic applications.
 CC AAP cDNA is also useful in gene therapy. The invention also relates to
 CC a method for screening a tissue sample for tumourigenic potential. AAP
 CC proteins are used to screen drugs or compounds that modulate AAP activity
 CC or expression as well as treating disorders characterised by insufficient
 CC or excessive production of AAP or production of AAP forms that have
 CC decreased or aberrant activity compared to the wild type protein, or
 CC modulate biological function that involve AAP. The present sequence
 CC is human BAZF (hBAZF) protein which is an angiogenesis associated
 CC protein (AAP) of the invention. BAZF is a Bcl-6 (L2A3) homolog, a
 CC transcription repressor that controls germinal center formation and the
 CC T-cell dependent immune response. Human BAZF plays a role in cell
 CC proliferation.
 CC
 XX
 SQ Sequence 518 AA,

Query Match 100.0%; Score 2789; DB 22; Length 518;
 Best Local Similarity 100.0%; Pred. No. 6.8e-189;
 Matches 518; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGSPAPBEGALGYRETRHSSDVLGNLRLRGLITDVTLLVGGQPLRAKAVLIACS 60
 DB 1 MGSPAPBEGALGYRETRHSSDVLGNLRLRGLITDVTLLVGGQPLRAKAVLIACS 60
 QY 61 GFYYSIFRGAGVGVDSLPGGPEARGFAPLDPMYTSRLRSPATAPVLAATYLLQ 120
 DB 61 GFYYSIFRGAGVGVDSLPGGPEARGFAPLDPMYTSRLRSPATAPVLAATYLLQ 120
 QY 121 EHVVOACHRFIOASYEPLGISLRPLEAPPTPTAPPPGSRSEGHDPPTESRSCSQ 180
 DB 121 EHVVOACHRFIOASYEPLGISLRPLEAPPTPTAPPPGSRSEGHDPPTESRSCSQ 180
 QY 181 PPSBPASDPYACMKKKYIVLNSQASQSLVGRSSGQPCQARLPSCDEASSSSSS 240
 DB 181 PPSBPASDPYACMKKKYIVLNSQASQSLVGRSSGQPCQARLPSCDEASSSSSS 240
 QY 241 SSSSSSEGFIPGOSRLSPATVOPFCGAPASTPYLLTSQADTSSPERARPIPGSE 300
 DB 241 SSSSSSEGFIPGOSRLSPATVOPFCGAPASTPYLLTSQADTSSPERARPIPGSE 300
 QY 301 FFSQONCEAVAGSSGLDSLVPGEDEKPYKQLCRSSFRYKGMASHRTVHTGEKPYHCS 360
 DB 301 FFSQONCEAVAGSSGLDSLVPGEDEKPYKQLCRSSFRYKGMASHRTVHTGEKPYHCS 360
 QY 361 ICGARFRPANKLTHSRHISGEKPYKCTGSRFVQVRSQPSGFGQKPARGVGQKGF 420
 DB 361 ICGARFRPANKLTHSRHISGEKPYKCTGSRFVQVRSQPSGFGQKPARGVGQKGF 420
 QY 421 CSSGORDLSRPSQVHLRAHVLIHGEKPYPCPTGCTGRFRHLQTLKSHVRIHTGKPYH 480
 DB 421 CSSGORDLSRPSQVHLRAHVLIHGEKPYPCPTGCTGRFRHLQTLKSHVRIHTGKPYH 480
 QY 481 CDPCGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518
 DB 481 CDPCGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518
 QY 481 CDPCGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518
 DB 481 CDPCGLHFRHKSQRLRLRQKHGAATNTKVAHYILGSP 518

RESULT 2

ABB57289.
 ID ABB57289 standard; Protein; 707 AA.

AC ABB57289;

DT 07-MAR-2002 (first entry)

DE Mouse ischaemic condition related protein sequence SEQ ID NO:814.

KW Mouse; ischaemia; compressive ischaemia; occlusive ischaemia;
 KW vasoepatic ischaemia; ischaemic condition; ischaemic disease.

OS Mus musculus.

PN MO200188188-A2.

PD 22-NOV-2001.

PF 18-MAY-2001; 2001WO-0P04192.

PR 18-MAY-2000; 2000JP-0145977.

PA (UNYI-) UNIV NIHON SCHOOL JURIDICAL PERSON.

PI Ishikawa K, Asai S, Takahashi Y, Nagata T, Ishii Y;

DR WPI; 2002-034733/04.

DR N-PSDB; AB199745.

PT Examining the ischaemic condition (e.g. occlusive ischaemia) by measuring
 PT expression levels of particular genes defined in the specification or
 PT by determining the expression profile of a gene group comprising these
 PT genes -

PS Claim 2, Page 2004-2007; 2630pp; English.

XX The present invention describes a method for examining ischaemic
 CC conditions, comprising measuring the expression levels of particular
 CC genes (I) in a test sample or determining the expression profile of a
 CC gene group in the sample comprising genes selected from (I). The method
 CC is useful for examining the ischaemic condition (e.g. compressive
 CC ischaemia, occlusive ischaemia or vasoepatic ischaemia) by measuring
 CC expression levels of particular genes (AB199202 to AB199912, encoding
 CC the protein sequences in ABB57020 to ABB57374) or by determining the
 CC expression profile of a gene group comprising these genes. The
 CC expression profile of a gene group comprising these genes are
 CC used as an indicator when screening for ischaemic condition-improving
 CC drugs or therapeutics for ischaemic diseases. AB199913 and AB199914
 CC represent PCR primers for a mouse ischaemic condition related sequence,
 CC which are used in the exemplification of the present invention.

SQ Sequence 707 AA,

Query Match 40.8%; Score 1138; DB 23; Length 707;
 Best Local Similarity 36.4%; Pred. No. 4e-72;
 Matches 269; Conservative 51; Mismatches 149; Indels 270; Gaps 14;

QY 1 MGSPAPBEGALGYRETRHSSDVLGNLRLRGLITDVTLLVGGQPLRAKAVLIACS 60
 DB 1 MASP-DSCL---QFTRHSDVLLNRLRSRLITDVTLVVSRBOFRAKTVLMCS 54
 QY 61 GFYYSIFRGAGVGVDSLPGGPEARGFAPLDPMYTSRLRSPATAPVLAATYLLQ 120
 DB 61 GFYYSIFRGAGVGVDSLPGGPEARGFAPLDPMYTSRLRSPATAPVLAATYLLQ 120
 QY 121 EHVVOACHRFIOASYEPLGISLRPLEA----- 148
 DB 121 EHVVOACHRFIOASYEPLGISLRPLEA----- 148
 QY 149 -----PT----- 151
 DB 175 CESRAPAPLYSGLSTPPASYPMYSHLPLSTFLFSDEBLRDAPRMPVANPPKRALPCD 234

```

QY 152 ----- 151
DB 235 SAROVNEYSRPAMEVSPSLCHSNISPYKAEVBEARSIDHVSVEPGKPAVPSANAPY 294
QY 152 ----- 176
DB 295 PCDKASKEERPSSEDEIALHPEPNAPLNRRKGLVSPQSPKSDQCPNPTSCSSKNA 354
QY 177 ----- 225
DB 355 CILQASGSPAKSPDPKACNMKKYKIVLNS-----QASQAGSLVGRSSGQPCPOA 225
QY 226 RLPSPGDEAS--SSSSSSSSSSSEEGPIPGPOSRL-----SPTATVQF----- 266
DB 415 CQPMPEPAILDQSPFKLSASGEDSTIP--QASRLNLLVNRSLGSPRSSSSSHSPLYMHP 473
QY 267 -KCGAPASTPYLLTQAODTSGSPSERAPLPSESE-----FSCQNCENAVGCCSG 316
DB 474 PKCTSCGSGSPQHTKCHTACPTPEEMGETOSEYSDSSCENGTFPCNECCDRFSEAS 533
QY 317 LD-SLVPGDEDEKPYKQLCRSSFRYKGNLASHRTVTGGEKPYHCSICGARFNRPAVLKTH 375
DB 534 LGRHTLQTHSDKPYKCDRCQASFRYKGNLASHKTVTGGEKPYRCNICGAQFNRPAVLKTH 593
QY 376 SRHSGEKPKCETCGSRFVQVRSPSGFQGRPARGVGQKGFSSQRODLKSPPSQV 435
DB 594 TRHSGEKPKCETCGARFV-----QV 615
QY 436 AHLRAHVLHTGKPYPCPTCCTFRHLQTLKSHVRIHTGKPYHCDPCGLHFRKSOQR 495
DB 616 AHLRAHVLHTGKPYPCETCCTFRHLQTLKSHLRIHTGKPYHCEKCNLHFRKSOQR 675
QY 496 LHLRQKGAATNTKVHYHI 514
DB 676 LHLRQKGAATNTKVQYRV 694

RESULT 3
AA78793
ID AA78793 standard; protein; 706 AA.
XX AA78793;
XX
XX 19-MAY-2000 (first entry)
XX
XX Human BCL-6 protein sequence.
XX
XX Bcl-6; human; B-cell lymphoma; regulator; non-Hodgkin's lymphoma;
XX diffuse type B-cell lymphoma.
XX
XX Homo sapiens.
XX
XX WO200000185-A1.
XX
XX 06-JAN-2000.
XX
XX 30-JUN-1999; 99WO-US14703.
XX
XX 30-JUN-1998; 98US-0107058.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX Dalla-Favera R, Niu H,
XX
XX WPI; 2000-160631/14.
XX
XX Novel methods for regulating BCL-6 levels in cells used to treat humans
XX with lymphoma
XX
XX Example 2; Fig 10; 159pp; English.
XX
XX This sequence represents the human bcl-6 protein sequence. The invention
XX relates to a vertebrate bcl-6 locus which is the breakpoint cluster

```

region in B-cell lymphomas, and containing a bcl-6 gene encoding a BCL-6 polypeptide. Administration of a molecule which induces phosphorylation of BCL-6 and thereby induces BCL-6 degradation, can be used as a method of regulating BCL-6 in cells. The method of the invention can be used to regulate, and especially to decrease BCL-6 levels in cells. The method may also be used to screen putative therapeutic agents for treatment of non-Hodgkin's lymphoma, by contacting cells from lymphoma and normal cells with the agent, and after a period of time comparing the amount of bcl-6 nucleic acid in each sample, a difference indicating the effectiveness of the agent. The bcl-6 gene is a source of probes and primers, which are used to diagnose diffuse-type B cell lymphoma and B cell lymphoma in a subject. Anti-BCL-6 antibodies may also be used for this purpose. The methods are useful for treating non-Hodgkin's lymphoma.

XX Sequence 706 AA;

Query Match 40.2%; Score 1122.5; DB 21; Length 706;
Best Local Similarity 36.3%; Pred. No. Se-71;
Matches 268; Conservative 58; Mismatches 142; Indels 271; Gaps 17;

```

QY 1 MGSAPAPGALGVBEFTRHSDVGNINELRLGILTDVTLVVGQPLRAKAVLIACS 60
DB 1 MASPA--DSCT-----QFTRHASDVLLNLRSLRILDVIVVSRBQFRANKTYLMACS 54
QY 61 GPFYSIFGRAGVGVVLSLPGEPARGAPLLDMYTSRLSLSPATAPAVLAATYLLQM 120
DB 55 GLFYSIFPDQKCNLSVINLDEINPEGFCLIDPMYTSRLNLRGNTMAVATMYLLQM 114
QY 121 EHVVAQCHRFIOAS-----YEPGIS----- 134
DB 115 EHVVDTCRFKFIKASEAENVSAIKPREBEFLNSRMLPMODIMAYRGRVVENNLPLRSABG 174
QY 135 ----- 141
DB 175 CSRAFAPSLVSGLSTPPASYSMYSHLVSLSLBSDEFRDYMVAAPFPKERRALPCDS 234
QY 142 -----LRL----- 145
DB 235 ARPVGESRPTLBEVSPNVCHSNISYPKETIPBEARSDMHWASVAGLKAAPASANAPY 294
QY 146 -----EAEPPT-----PPTAP-----PGSPRSSEGHFPDPTSRG----- 176
DB 295 PCDKASKEERPSSEDEIALHPEPNAPLNRRKGLVSPQSPKSDQCPNPTSCSSKNA 354
QY 177 -----CSQPPSPASPPDKACNMKKYKIVLNS-----QASQAGSLVGRSSGQPCP 223
DB 355 ILQASGSPAKSPDPKACNMKKYKIVLNSLNQNAKPGEGEQAELRLSPRAYTAPAC 414
QY 224 QARLPSPGDEASSSSSSSSSEEGPIPGPOSRL-----SPTATVQF----- 266
DB 415 QPPM--EPENLDLQSPFKLSASGEDSTIP--QASRLNLLVNRSLGSPRSSSSSHSPLYMHP 472
QY 267 ---KCG--APASTPYLLTQ---AODTSGSPSERAPLPSESEFSCQNCENAVGCCSG 316
DB 473 PKCTSCGSGSPQHAEMCLHTAGPTFAEEMGETOSEYSDSSCENGTFPCNECCDRFSEAS 532
QY 317 LD-SLVPGDEDEKPYKQLCRSSFRYKGNLASHRTVTGGEKPYHCSICGARFNRPAVLKTH 375
DB 533 LGRHTLQTHSDKPYKCDRCQASFRYKGNLASHKTVTGGEKPYRCNICGAQFNRPAVLKTH 592
QY 376 SRHSGEKPKCETCGSRFVQVRSPSGFQGRPARGVGQKGFSSQRODLKSPPSQV 435
DB 593 TRHSGEKPKCETCGARFV-----QV 614
QY 436 AHLRAHVLHTGKPYPCPTCCTFRHLQTLKSHVRIHTGKPYHCDPCGLHFRKSOQR 495
DB 615 AHLRAHVLHTGKPYPCETCCTFRHLQTLKSHLRIHTGKPYHCEKCNLHFRKSOQR 674
QY 496 LHLRQKGAATNTKVHYHI 514
DB 675 LHLRQKGAATNTKVQYRV 693

```


RESULT 4
 AAB29640
 ID AAB29640 standard; Protein; 706 AA.
 XX
 AC AAB29640;
 XX
 DT 23-FEB-2001 (first entry)
 XX
 DE Human bcl-6 transcriptional repressor.
 XX
 KM Human; bcl-6; transcriptional repressor; germinal centre formation;
 KM Th-2 mediated antibody affinity maturation; apoptosis regulator;
 KM chromosome 3q27; lymphoma; acute lymphoblastic leukaemia;
 KM post-transplant lymphoproliferative disorder; expression inhibition;
 KM anticancer therapy.
 XX
 OS Homo sapiens.
 XX
 PN US6140125-A.
 XX
 PD 31-OCT-2000.
 XX
 PF 15-OCT-1999; 99US-0418640.
 XX
 PR 15-OCT-1999; 99US-0418640.
 XX
 PA (ISIS-) ISIS PHARM INC.
 XX
 PI Taylor JK, Cowseert LM,
 XX
 DR WPI; 2001-048959/06.
 DR N-PSDB; AAC81137.
 XX
 PT Anticancer compounds which specifically hybridize with and inhibit human
 PT bcl-6 expression, useful for treating bcl-6 related disorders, and
 PT preventing or delaying inflammation or tumor formation
 XX
 PS Disclosure; Column 47-52; 42pp; English.
 XX
 CC This sequence represents human bcl-6. Bcl-6 (also known as B-cell
 CC CLL/lymphoma 6, zinc finger protein 51 and LZ23) is a sequence-
 CC specific DNA-binding transcriptional repressor. The bcl-6 gene is
 CC expressed in germinal centre B- and T-cells and is required for
 CC germinal centre formation and Th-2 mediated antibody affinity maturation.
 CC Bcl-6 may also play a role in the regulation of apoptosis. The bcl-6 gene
 CC is located on chromosome 3q27, a region which undergoes a high frequency
 CC of translocation events. Such chromosomal translocations can result in
 CC aberrant forms of bcl-6, which are strongly implicated in the
 CC pathogenesis of several types of lymphoma, and have also been reported
 CC in acute lymphoblastic leukaemia and post-transplant lymphoproliferative
 CC disorders. The invention relates to antisense oligonucleotides targeted
 CC to the human bcl-6 gene, which inhibit its expression. A series of
 CC oligonucleotides (AAC8114-C8123) were designed to target different
 CC regions of the human bcl-6 mRNA, and were analysed for their effect on
 CC bcl-6 mRNA levels by quantitative real-time PCR. The oligonucleotides of
 CC the invention are useful for diagnosis, prevention and treatment of
 CC conditions associated with aberrant forms of bcl-6, such as lymphomas,
 CC acute lymphoblastic leukaemia and post-transplant lymphoproliferative
 CC disorders.
 CC
 SQ Sequence 706 AA;
 XX
 Query Match 40.2%; Score 1122.5; DB 22; Length 706;
 Best Local Similarity 36.3%; Pred. No. 5e-71;
 Matches 268; Conservative 58; Mismatches 142; Indels 271; Gaps 17;
 QY 1 MGSPAAEGALGYRETRHSSDVLGNLRLGILTDVTLVVGQPLRAHRAVLIACS 60
 DB 1 MASPA--DSCI-----GTRHSDVLTNLRSLRDLITDVIVVSEQPRAHKTVLMACS 54
 QY 61 GPFYISIFRGAGVGVVLSLPGGEARGFAPLDDFWYTSRLSPATPAVLAAYTLQW 120
 DB 55 GLFYISIFTDQLKCNLSVINLDPINPEGFCILDDFWYTSRLNREGNIMAVMATAYLQW 114

QY 121 EHVQACHRFIOAS----- 134
 DB 115 EHVDTCKRFIKASEEMWSAIKPREBEFLNSRMLMPQIMAYRGREVNNLPLRSARG 174
 QY 135 -----YEP LGIS----- 141
 DB 175 CESRAFAPSLYGLSTPPASYSMSHLPVSSILFSDDEFRDVMVAMPYKRALPCDS 234
 QY 142 -----LRPL----- 145
 DB 235 ARPVEYSRPLEVSPNVCNSNIYSPKTTIEBARSMDHVSVAQLKAAASARNAPYF 294
 QY 146 -----EAEPPT-----PPTAP-----PPGSPRSEGHDPPTESRS----- 176
 DB 295 PCDKASKEEERSSEDEIALHFEPPNAPLNKRGVSPQSPQSDQPNPTACSSKXAC 354
 QY 177 ---CSQGPSPASPDPKACNMKKRYIVINS-----QASQAGSLVGRSSGQPCP 223
 DB 355 ILQASGSPPAKSPDPKACNMKKRYFIVLNSINQNAKPGGPGQAEIGRLSPRAYAPAC 414
 QY 224 QARLPBGDEASSSSSSSSSEBGPDPGOSRL-----SPTATVQF----- 266
 DB 415 QPPM-BPENLDIQSPKUSASGEDSTIP-QASRLNNIVNRSMTGSPRSSHSPLYMHP 472
 QY 267 ---KCG--APASTPYLLTSQ-----AODTSGSPERAPLPSEFFPSCQNCBAVAGCSG 316
 DB 473 PKCTGCGSOSPQHAEMCLHTAGPTFAEMGETOSHSOSSCENGAFPCNECDRCRSEELS 532
 QY 317 LD-SLVPGDEDPYKQQLCRSSFRYKGNLASHRTYHTGKPYHCSTCGAREFRPANKTH 375
 DB 533 LKRHTLQTHSDKPYKCDRCQASFRYKGNLASHRTYHTGKPYRCNICGQAFNRPANKTH 592
 QY 376 SRIHSEKPYKCTGSPRVQVRSQPPSGFGKPARCGVGQKGRGSSQRDLKSPSQV 435
 DB 593 TIIHSEKPYKCTGCGARV-----OV 614
 QY 436 AILRAVLIHTGEKPYPCPTGTRFRHLQTLKSHVRIHTGEKPYHCDPCGLHFRHKSQAR 495
 DB 615 AILRAVLIHTGEKPYPCGTRFRHLQTLKSHRIHTGEKPYHCEKCNLHFRHKSQAR 674
 QY 496 LHLRQHGANTYKQVYHI 514
 DB 675 LHLRQHGAIYNTYKQYRV 693
 RESULT 5
 AAR68743
 ID AAR68743 standard; Protein; 706 AA.
 AC AAR68743;
 XX
 DT 19-UTL-1995 (first entry)
 XX
 DE BCL-6 zinc finger protein.
 XX
 KM bcl-6 locus; non-Hodgkin lymphoma; B-cell lymphoma; B-lymphocyte;
 KM diagnostic; therapeutic; chromosome-3q27; translocation;
 KM proto-oncogene; diffuse large cell lymphoma; DLCL; zinc finger.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Region 520..541
 FT /label= Zinc-finger
 FT Region 548..568
 FT /label= Zinc-finger
 FT Region 576..596
 FT /label= Zinc-finger
 FT Region 604..624
 FT /label= Zinc-finger
 FT Region 632..652
 FT /label= Zinc-finger

```

FT Region 660..681
FT /label= Zinc-finger
XX MO9429343-A.
XX
XX 22-DEC-1994.
XX
XX 09-JUN-1994; 94WO-US06669.
XX
XX 09-JUN-1993; 93US-0074967.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX (SLOK ) SLOAN KETTERING INST CANCER.
XX
XX Chaganti RK, Dalla-Favera R, Chaganti RS;
XX
XX WPI, 1995-036403/05.
XX N-PSDB; AAQ68743.
XX
XX Nucleic acid from genetic locus bcl-6 - used to develop probe.
XX for diagnosis and therapy of B-cell lymphoma and non-Hodgkin's
XX lymphoma
XX
XX Disclosure; Page 90-94; 129pp; English.
XX
XX DNA was extd. from tumor tissue of 2 cases of IgM-producing diffuse-
XX type B-cell NHL carrying the t(3;14)(q27;q32) translocation. DNA
XX analysis showed that the breakpoints on 3q27 were located within 3
XX kb of the same genomic locus, which was designated bcl-6. The
XX human bcl-6 locus was identified by screening a phage cDNA library
XX constructed from Bjab B-cell lymphoma mRNA. The zinc finger
XX protein encoded by bcl-6 is given in AAR68743. ECL-6 is a
XX proto-oncogene specifically involved in the pathogenesis of diffuse
XX large cell lymphoma.
XX
XX Sequence 706 AA:
SQ
Query Match 39.3%; Score 1095.5; DB 16; Length 706;
Best Local Similarity 35.7%; Pred. No. 4e-69;
Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;
QY 1 MGSPAAEGALGVREFTTHSSDVLTGNLNLRLGILITDTVTLVGGQPLRAHKAVALIACS 60
DB 1 MASPA--DECI-----QFTHARDVLTNLRLSRDITDVTIVSRSEGFRAKTVLMAMR 54
QY 61 GFYGSIFRGRAGVGVVLSLPGEPARGAPLIDPFYTRSLKSPATAVAVLAATYLOM 120
DB 55 GLFYSLFTDQLKCNLSVINLDEPEINDEGFCILDFMYTSRLNLREGNINAVATMYLOM 114
QY 121 EHVVOACHRFIONS-----YEPLGIS----- 134
DB 115 EHVVDTCRKFTKQSEAMVSAIKPPEEFNLNRMMPDIMAARGEVENNLPLRSAG 174
QY 135 ----- 141
DB 175 CESRAPEPLYSGLSTPPASYSWYSHLPVSSLLFSDPEFRDVMVPAVNPFRKRALPCDS 234
QY 142 ----- 145
DB 235 ARPVGESVPTLEVPNVCHSNIVSPKTEPBEARSDMHVSAEGIKPAASARNAVYF 294
QY 146 -----BAEPPT-----PPTAP-----PGSPRSEGHDPDPPTES---RSC 177
DB 295 PCDKASKEERFSSSEBIALHPEPPVAPLARKGLVSPQSDCCPNSPTEKCSKXAC 354
QY 178 -----SGPPSPAPDPKACNMKKYKIVINS-----QASQAGSLVGERSSGQPCP 223
DB 355 ILQSGSPPAKSPPTDEPKACSMKKYKIVINSLNQNAKPGQBEAGELGRSLSPRAYTAPAC 414
QY 224 QARLPBGCDASSSSSSSSSSSEEGPIPGQSL-----STATAVQP----- 266
DB 415 QPPM-EPENLDLQSPYKLSASGSDSTIP-QASRLNINIVNR3MTGSPRSSSESHSLYMP 472

```

```

QY 267 ----KCG--APASTPYLLTSQ-----ADPTSGSPSERARPLGSEPFSSCONCEAVAGCSSG 316
DB 473 PKCTSCGSGSPQAHMCLHTAGPTFAEMGSETQSEYSSSCENGAFNFNCEDCRSREAS 532
QY 317 LD-SLVPDEDEKPYKQCLCRSSFRYKGNLASHRYVHTGEKPYHSGICGARFRPRLNLT 375
DB 533 LKRHTLQTHSDKPYKCDRCQASFRYKGNLASHRYVHTGEKPYRNCI CAAQFRPRLNLT 592
QY 376 SRHSGEKPYKCECTGSRFVGVRSQPPSGFGQKPARGVGQKGGFCSSQRODLKSPFSGV 435
DB 593 TRHSGEKPYKCEITGANFV----- 614
QY 436 AHLRAHVLIHTGEKPYKCEPTGTRFRHLQTLKSHVRIHTGEKPYHCDPGLHFRKSQLR 495
DB 615 AHLRAHVLIHTGEKPYKCEICGTRFRHLQTLKSHVRIHTGEKPYHCBKCNHFRKSQLR 674
QY 496 LHLRQKGAATNTKVNHI 514
DB 675 LHLRQKGAITNTKVQYRV 693
XX
XX RESULT 6
XX AA78792
XX ID AA78792 standard; Protein; 706 AA.
XX
XX AA78792;
XX
XX 19-MAY-2000 (first entry)
XX
XX Human BCL-6 protein sequence.
XX
XX Bcl-6; human; B-cell lymphoma; regulator; non-Hodgkin's lymphoma;
XX diffuse type B-cell lymphoma.
XX
XX Homo sapiens.
XX
XX WO20000185-A1.
XX
XX 06-JAN-2000.
XX
XX 30-JUN-1999; 99WO-US14703.
XX
XX 30-JUN-1998; 98US-0107058.
XX
XX (UYCO ) UNIV COLUMBIA NEW YORK.
XX
XX Dalla-Favera R, Niu H;
XX
XX WPI; 2000-160631/14.
XX
XX N-PSDB; AA290110.
XX
XX Novel methods for regulating BCL-6 levels in cells used to treat humans
XX with lymphoma -
XX
XX Examples; Fig 9; 159pp; English.
XX
XX This sequence represents the human bcl-6 protein sequence. The invention
XX relates to a vertebrate bcl-6 locus which is the breakpoint cluster
XX region in B-cell lymphomas, and contains a bcl-6 gene encoding a BCL-6
XX polypeptide. Administration of a molecule which induces phosphorylation
XX of BCL-6 and thereby induces BCL-6 degradation, can be used as a method
XX of regulating BCL-6 in cells. The methods of the invention can be used to
XX regulate, and especially to decrease BCL-6 levels in cells. The method
XX may also be used to screen putative therapeutic agents for treatment of
XX non-Hodgkin's lymphoma, by contacting cells from lymphoma and normal
XX cells with the agent, and after a period of time comparing the amount of
XX bcl-6 nucleic acid in each sample, a difference indicating the
XX effectiveness of the agent. The bcl-6 gene is a source of probes and
XX primers, which are used to diagnose diffuse-type B cell lymphoma and B
XX cell lymphoma in a subject. Anti-BCL-6 antibodies may also be used for
XX this purpose. The methods are useful for treating non-Hodgkin's lymphoma.
XX

```

SQ Sequence 706 AA;
 Query Match 39.3%; Score 1095.5; DB 21; Length 706;
 Best Local Similarity 35.7%; Pred. No. 4e-69;
 Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;

```

QY 1 MGSPAAEGALGYVEFTRHSSDVLGNLNLRLGILITDVTLLVGGQPLRAHKAVALIACS 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MASPA--DSCI-----QTRHARDVLLNLNLRSDRLTDVIVVSRQFRAHKTIVLMAWR 54
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 GPFYSLFRGAGVGVLDLSLPGCEARGFAPLDFMTSTSLRLSPATPAVLAATYLM 120
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 55 GLFYSLFTDQKCNLYINLDPINPEGFCILLDFMTSTSLRLREGNIMAVMATAYLM 114
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 121 EHYVQACHRFQAS-----YEPGIS----- 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 115 EHYVDTCRKFIKASAEAMVSAIKPREELNSRMIMQDIMAYRGREVENNPLIASAPG 174
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 135 ----- 141
DB 175 CESRAFAPSLYGLSTPPASYSMYSHLPVSSLFSDDEFDMVMPVANPEKRALPCDS 234
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 142 ----- 145
DB 235 ARPVGESRPTLEVPNCHSNISPKETIPPEARSDMYSVAEGLKPAAPARAPAYF 294
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 146 -----EAPPT-----PPTAP-----PPGSRREGHPDPTES-----RSC 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 295 PCOKASKEERPSSEDEIALHFEPPNAPLRKGLVBPQSKDCQPNSTELCSKMC 354
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 178 ----SQQPSASPDPPKACNWKYKYIVLNS-----QASQAGSLVERSSQOPC 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 355 ILGGSGSPPAKSPDTPKACSMKKYKFIVLNLSQNAKPGPEQAEIGRLSPRAYTAPAC 414
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 224 QALPLPGDEMASSSSSSSSSEGPFGQSR-----SPRAIVQF----- 266
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 415 QPFW-EPENDLQSPYTLASGSDSTIP-QASRLNINIVNSMTGSPSSSESHSPLYMP 472
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 267 ----KCG--APASTPYLLTSQ-----AQTGSPSERARPLPGSEFSCQCEAVAGCSG 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 473 PKTSCSGSPQIAEMKLTHTAGPTFAEMERTSEVSDSCENGAFPCNECDRFEEMS 532
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 317 LD-SLVPGDEKPKYKCOLCRSSFRYKGNLASHRTVTGKPYKCSICGARFNPANLKT 375
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 533 LKSHTLQTHSDKPKYKCDRCQASFRYKGNLASHRTVTGKPYKCNICGAQFNPANLKT 592
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 376 SRHSGEKPCKCTCGSRFYQVRSQPPSGFGKPRAGGVQKGFCSGQKDLSPSPQY 435
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 593 TRHSGEKPCKCTCGARFV-----OV 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 436 AHLRAVLITHTGEKPYPCPTCGTRFRLQTLKSHVRLHTGEKPYHCDPGLAFRHKSQ 495
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 615 AHLRAVLITHTGEKPYPCPTCGTRFRLQTLKSHVRLHTGEKPYHCKCNLHRHKSQ 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 496 LHLRQKGAATNTKVAHYI 514
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 675 LHLRQKGAITNTKVQYRV 693
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
AAV73351 standard; Protein; 810 AA.
XX
XX
AC AAV73351;
XX
DT 24-FEB-2000 (first entry)
XX
DE HTRM clone 1484257 protein sequence.
XX
XX HTRM; human transcriptional regulatory molecule; arteriosclerosis; AIDS;
KM arteriosclerosis; cirrhosis; cancer; leukemia; diabetes mellitus;
KM Addison's disease; multiple sclerosis; rheumatoid arthritis; infection;
KM trauma; myasthenia gravis; adenocarcinoma; immune disorder; treatment.
  
```

XX OS Homo sapiens.
 XX PN MO957144-A2.
 XX PD 11-NOV-1999.
 XX PF 04-MAY-1999; 99WO-US09935.
 XX PR 05-MAY-1998; 98US-0084254.
 XX PR 07-AUG-1998; 98US-0095827.
 XX PR 02-OCT-1998; 98US-0102745.
 XX PA (INCY-) INCYTE PHARM INC.
 XX PI Hillman JL, Bandman O, Lai P, Yue H, Reddy R, Tang YT;
 XX PI Gerstein EH, Patterson C, Baughn MR, Azimzai Y, Lu DM;
 XX DR WPI; 2000-052941/04.
 XX DR N-PSDB; AA52436.
 XX PT New peptides useful for diagnosis, prevention and treatment of cancer
 XX PT and immune disorders.
 XX PS Claim 1; Page 112-114; 193pp; English.
 XX CC AAV73325-Y73389 are human transcriptional regulator molecule (HTRM)
 CC protein sequences. The HTRM protein and nucleotide sequences are useful
 CC for preventing or treating disorders associated with decreased expression
 CC or activity of HTRM which include cell proliferative disorders such as
 CC arteriosclerosis and cirrhosis; cancers including adenocarcinoma and
 CC leukemia; immune disorders such as AIDS, Addison's disease, diabetes
 CC mellitus, rheumatoid arthritis, multiple sclerosis, systemic lupus
 CC erythematosus, and myasthenia gravis; infections and trauma. Antagonists
 CC of the HTRM polypeptides are useful for treating or preventing disorders
 CC associated with increased expression or activity of HTRMs. HTRM
 CC polypeptides, their immunogenic fragments or oligopeptides are useful
 CC for screening libraries of compounds in drug screening techniques.
 CC Polynucleotides encoding HTRM are useful for blocking the transcription
 CC of mRNA and regulating gene function by modulating the activity of
 CC HTRM. Vectors expressing HTRM or agonists can also be used to prevent or
 CC treat disorder associated with decreased HTRM expression. Antibodies
 CC which specifically bind HTRM and polynucleotides encoding HTRM are useful
 CC for diagnosing disorders associated with the expression of HTRM,
 CC particularly in assays that detect the expression of HTRM. Nucleotide
 CC sequences encoding HTRM may be useful to generate hybridization probes
 CC useful in mapping the naturally occurring genomic sequence and to detect
 CC differences in gene sequences among normal, carrier and affected
 CC individuals. Using diagnostic assays, cancer can be detected prior to
 CC the appearance of clinical symptoms and thereby progression of cancer can
 CC be prevented by aggressive treatment or preventive measures.

SQ Sequence 810 AA;
 Query Match 20.9%; Score 582; DB 21; Length 810;
 Best Local Similarity 27.3%; Pred. No. 8.8e-33;
 Matches 175; Conservative 69; Mismatches 182; Indels 216; Gaps 21;

```

QY 16 EFTRHSDVLGNLNLRLGILITDVTLLVGGQPLRAHKAVALIACGFFSIRGAG-VG 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2 DPHQSHVLEQNLNOORQLGCLDCTFVVDGPHFAHRAVLAACEYFMLEVVDKDVH 61
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 75 VDVLSPGPEARGFAPLDFMTYRSLRSLSPATPAVLAATYLMENHVQACHRFQAS 134
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 62 LDI-----SNAAGQVLEFMTYKLSLSPENVDVLAATYLMODIITKCH----- 109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 135 YEPLGISLPLEAEPTP-----TAPPGSPRSE-----GH 167
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 110 -----ALKSL-AEATSPGCVGAALAQKVCVPSGQKAKAEKVATSTLSRLQGR 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 168 PPPPEPSR-----SCSQG-----PPSPASDPKACNWKYKYIVL 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 163 STPIGSRDLKEERGQASASGAEOTEKADAPRAPPVVELKPDPTS-----G 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
  
```


XX 26-JUL-2001.
 PD
 XX 26-DEC-2000; 2000MO-US34263.
 PF
 XX 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA
 XX (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR MPI; 2001-442253/47.
 DR N-PSDB; AAI58428.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 4; SEQ ID NO 2417; 10078bp; English.
 PS
 XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA038642-AA042213) with nootropic,
 CC immunosuppressant and cytoskeletal activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.
 CC
 SQ Sequence 803 AA;
 Query Match 20.6%; Score 574.5; DB 22; Length 803;
 Best Local Similarity 29.4%; Pred. No. 3e-32;
 Matches 180; Conservative 64; Mismatches 205; Indels 163; Gaps 20;

DB 280 LGSEARGLSGTGYGRTESKAYGVYHKCEDCKEFTHTGNFKRIHITGKEPSCREB 339
 QY 308 EAV---AGCSSGLDVLVPGDEDEKPKYKCOLCRSPRY----- 340
 DB 340 SKAFSDPPACKNAHEKTHSP--LKPYGCEBCKSLRLISLNLHKKRHSGEARYCEDCG 396
 QY 341 ----KGNLASHRTVHTGKPYHCISICG-----ARPN 367
 DB 397 KLFITSGNLKRLQLVHSGEKPYQCDYCGRSFSDPTSKMRHLETHDTEHKCPCDKKFN 456
 QY 368 RPAHLKTHSRHNSGKPYKCFCTGSRFPVYRSQPPSSGQKARAGVGQKGFCCSQROD 427
 DB 457 QVGNLKAHLKTHIADGPKRCRCGRQFT-----TSGNLKRLRLHSGEKPYVCIHCROQ 510
 QY 428 LKSPSSQVAHLRAVHLHTGKPYPCPTGTRFRHLQTLKSHVRHTGKPYHCDPCGLH 487
 DB 511 PADPCA---LGRVRIHTGKPYCCGWCCKAFTOASSLIHVRQHTGKPYVCRCGR 566
 QY 488 FRHKSQRLHLR 499
 DB 567 FVOSSQLANHIR 578
 RESULT 10
 ID AAM41058
 AAM41058 standard; Protein; 804 AA.
 AC AAM41058;
 DT 22-OCT-2001 (first entry)
 DE Human polypeptide SEQ ID NO 5989.
 XX
 KW Human; nootropic; immunosuppressant; cytoskeletal; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000MO-US34263.
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA
 XX (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Weinman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX
 DR MPI; 2001-442253/47.
 DR N-PSDB; AAI60214.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 XX
 PS Example 2; SEQ ID NO 5989; 10078bp; English.
 PS
 CC The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AA038642-AA042213) with nootropic,

CC Immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilization of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

CC Sequence 804 AA:

Query Match 20.6%; Score 574.5; DB 22; Length 804;
 Best Local Similarity 29.4%; Pred. No. 3e-32;
 Matches 180; Conservative 64; Mismatches 205; Indels 163; Gaps 20;

QY 16 EFTRRSSDVLGNLNLRLGILITDVTLLVGGQPLRAKAVLACSGFFYSIFRRAG-VG 74
 DB 3 DEPOHSQVLEBQUNQKQGLCDCTFVDVGHFKAKKAVLAACSEYFMFLFVDQDVVH 62
 QY 75 VDVLSLPGGEARGFAPLIDFMYTSRLSPATAPAVLAATVQMEHVQAQCHRFQAS 134
 DB 63 LDI-----SNAGAGQVLEFMYTAKLSIPENVDDVLAATFLQVQDITICH----- 110
 QY 135 YEPPLGISLAPLEAPPTP-----TAPPGSPRSEGHDPPTESRSCSQGPSPASPD 188
 DB 111 -----ALKSL-AEPATSPCGNAEALATGCGDRAKEEKVATSTLSRLQAGSTPIGP- 162
 QY 189 PRACMKKTKYIVLNSQASQASGLVGERSSGQPCQAPLP-----SGDEASSSSSSSSSS 245
 DB 163 --SRDLKERGGQASAAAGAEQTEKADAPREPPPELKPDPPTSGMAAEAALESSE 220
 QY 246 EEGPI-PGPQSRISPTATVQPKCA-PAST-----PYLLTSSQ-----QDTSGPS-- 290
 DB 221 QEMVEVPARKGEBOKEQEBEGAGPAVEBESQLNENGAPEMENEBESAGTDSQGE 280
 QY 291 --ERARPLP-----GSEF-----PSCQNC 307
 DB 281 LGSSEARGLSGTGDTESKAVGSVHKCEDCKEFTHTGTFKRIHITGKEPFCREB 340
 QY 308 EAV-----AGCSSGLDLSVGDDEKPKYKQLCRSSFRY----- 340
 DB 341 SKAFSDPAACKAEKTHSP--LKPYGCEBCKSKSYRLISLMLRKRHSGEARCEDCG 397
 QY 341 -----KGNLASHRTVTGGEKPYHCSICG-----ARPN 367
 DB 398 KLPFTSGNLKRIQVLHSGEKPYQCDYCGSFDPSTKMELETHTDIXHKCPCHCKKN 457
 QY 368 RPANLKTSHRISGEEKPYKCEGSRFVQVRSOPSPSGFCQKPARGVGQKGFCSQROD 427
 DB 458 QVGNLKAHKKIHADPDLKRCGKQFT-----TSGNLKRIHISGEEKPYVCICHQRO 511
 QY 428 LKSPSQVAVLRAHVLHITGEEKPYPCPTGTPRHQLKSHVRIITGGEKPYHCDPCGLH 487
 DB 512 FADPGA---LQNHVAIHITGEEKPYCQCMGKATQASSLIHVRQHTGGEKPYVCEGCRG 567
 QY 488 FRHKSQLRLHLR 499
 DB 568 FVQSSQLANHIR 579
 RESULT 11
 AAMA7790
 ID AAMA7790 standard; Protein; 610 AA.
 AC AAMA7790;
 XX
 DT 01-MAR-2002 (first entry)

XX Murine myoneurine.
 DE Murine; myoneurine; cytostatic; neurological; muscular; gene therapy;
 XX cancer; neuromuscular disease; dystrophy; polymyositis; rhabdomyosarcoma.
 OS Mus sp.
 PN FR2808535-A1.
 PD 09-NOV-2001.
 XX 05-MAY-2000; 2000FR-0005792.
 PF 05-MAY-2000; 2000FR-0005792.
 PR 05-MAY-2000; 2000FR-0005792.
 PA (ALLI/) ALLIEL P M.
 PI Seddiq N, Clifuentes DC, Rieger F, Perin JP;
 XX WPI; 2002-019508/03.
 DR
 PT New nucleic acid encoding human myoneurin, useful for diagnosis,
 PT prognosis and treatment of cancer and neurological diseases, comprises
 PT a polynucleotide encoding the human myoneurin protein -
 PS Claim 13; Page 61; 64pp; French.
 CC The present sequence is the protein sequence for murine myoneurin. The
 CC myoneurin coding sequence and its fragments are useful as primers and
 CC probes, for diagnosis, prognosis or assessing risk of human diseases,
 CC particularly cancers and neuromuscular diseases (e.g. dystrophy,
 CC polymyositis or rhabdomyosarcoma). The coding sequence and its fragments
 CC are also useful in gene therapy for these diseases.

XX Sequence 610 AA:

Query Match 17.3%; Score 483; DB 23; Length 610;
 Best Local Similarity 27.2%; Pred. No. 6.2e-26;
 Matches 152; Conservative 56; Mismatches 215; Indels 136; Gaps 14;

QY 16 EFTRRSSDVLGNLNLRLGILITDVTLLVGGQPLRAKAVLACSGFFYSIFRRAG--GV 73
 DB 2 QVSHHCEHLLEILNKQRBAGFLCDCTVIGSFQFYAHNVLASFSEYFGAIVRSTSENV 61
 QY 74 QVDVLSLPGGEARGFAPLIDFMYTSRLSPATAPAVLAATVQMEHVQAQ----- 127
 DB 62 FLDSQV---KADFQKLEIFLYTGLNLSWVKEIHQADYLVKEVVYKCKIKMED 117
 QY 128 HRFIQASYEPLGISLAPLEAPPTPTAPPPG-----PRSEGHDPPTESRSC 177
 DB 118 FAFI-----ABSSSTEISITGNTILNQOALCLTATRTNNREKSEVSDS 162
 QY 178 SQGPSPASPPEKAMKMKYKVIYVNSQASQASGLVGERSSGQPCQAPLPSPGDEASSSS 237
 DB 153 VQANPKPALTKKSSQSKKKKAFSSQKPG-----SKAVQPSDVLBSAVELFLD 214
 QY 238 SSSSSSSSEEGIFPPQS-----RLSTATV-----QPCG----- 269
 DB 235 TSKLSSPVEQIITQNDSELELTSVENTPEPTODIVTVVTKRRRKSQSHCALKEHSM 274
 QY 270 --APASTPYLLTSQAD-----TSGPSPRARPL-----PGSEPFSCON 306
 DB 275 SNIAVSKSPYLEMNGEFLDQRFSAKAMKMTCCGVNPSBASLRHMRHINQVRYVCHL 334
 QY 307 C-EAVAGSSGLDLSVGDDEKPKYKQLCRSSFRYKGNLASHRTVHTG-EKPYHCSICGA 364
 DB 335 CGKAFIQCNQKTHVTRHTGSRPYKCELDCKGFAQKQLVPHSRMHSEKPYKCDVCL 394
 QY 365 RPNRANLKTSHRISGEEKPYKCEGSRFVQVRSOPSPSGFCQKPARGVGQKGFCSQ 424
 DB 395 QPATSNLKIHAIRKRSGEKPYVCDRCGRF----- 424

QY 425 RODLKSPSPQVAHLRAHVLIIHTGEKPYCPTGTRFRHLITLTKSHVRIHTGEKPYHCDCP 484
 Db 425 -----AQAStLTIVVRHRTGTEKPYVCDTCKAFVSSSLITHSKRTGTEKPYICGIC 476
 QY 485 GLHFRHKSQRLHLRQKHG 503
 Db 477 GKSFSSGELNKHFRSHTG 495

RESULT 12
 AAM40035
 ID AAM40035 standard; Protein: 611 AA.
 AC AAM40035;
 XX 22-OCT-2001 (first entry).
 DE Human polypeptide SEQ ID NO 3180.
 XX
 XX Human; noctropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KM peripheral nervous system; neuropathy; central nervous system; CNS;
 KM Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KM chemokine; thrombolytic; drug screening; arthritis; inflammation;
 KM leukaemia.
 XX
 OS Homo sapiens.
 PN WO200153312-A1.
 XX 26-JUL-2001.
 PD 26-DEC-2000; 2000MO-US34263.
 PF 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 PA
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 XX WPI; 2001-442253/47.
 DR N-PSDB; AAI59191.
 XX
 PT Novel nucleic acids and polypeptides, useful for treating disorders
 such as central nervous system injuries -
 PS
 PS Example 4; SEQ ID NO 3180; 10078BP; English.
 XX
 XX The invention relates to human nucleic acids (AA157798-AA161369) and
 CC the encoded polypeptides (AAM38642-AAM42213) with noctropic, and
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localized neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: immune system suppression,
 CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemia and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX SQ Sequence 611 AA;
 Query Match 17.3%; Score 482.5; DB 22; Length 611;
 Best Local Similarity 26.9%; Pred. No. 6, 8e-26;
 Matches 151; Conservative 71; Mismatches 238; Indels 101; Gaps 17;

QY 16 EFTRRSHSDVGLNINELRLRGLITDVLVGGQPLRAHKAVALIACGFFYSIFRGRA--GV 73
 Db 2 QYSHHCEHLERLNNQREAGFLCDCTIYIGFQFAHNRVNLASFEGAIYRSTSENNV 61
 QY 74 GVDVLSLPGGPARGAPLLDPMYTSRLRLSPATPAVLAATYIQMEHYVOACH----- 128
 Db 62 PLDQSQV---KADFQKLEFIYTGTLNLSWNVKEHQADYKVEVYTKXIKMED 117
 QY 129 -----RFIOASIEPLGISLRP-----LEAPRPPTAPRPG 159
 Db 118 FAFIANPSTETISITGNTIELNQITCLTLRDYNNREKSEVSTDILQANPKGALAKSS 177
 QY 160 -SPRSRSGHPDPTESRSCSQP-----PSPASPPKACMKR---KY 197
 Db 178 QTKKKKKKAFNSPKTGQNTVQYPSDILENASVELFLDANKLPTPVVEQVAQINDNSELEL 237
 QY 198 KTIYVNSQASQ--AGSLVGERSSGQPCQARLP--GDEASSSSSSSSSSSEEGPIPG 252
 Db 238 TSVENTFPADIVHTVTVKRRGKSGPICALKSHSMNINASKSPYEAKENSGEE--LDQ 295
 QY 253 POSRLSPITATVQFCGAPASTFPVLLTSGAODTSGSPSERAR---LPGSFFSCQNC 308
 Db 296 RYSKAKPMQNT---CGKVFSEASSLRHRMRHKG---VAPYCHLCKGKFTCNOLK 346
 QY 309 AVAGCSSLGLSLVPDEDEPKPYKQQLCRSPFRYKYNLAGHRTVHTG--EKRYHSGICGAREN 367
 Db 347 THVRTHTG-----EKPYKCELCDGFAQKQVLVFRSHHGEKPYKCDVNLQGA 397
 QY 368 RPAANLKTSHRITSGEKPYKCECTGSRFVQVNSQPSGCGKARAGVGQKGFSGSROD 427
 Db 398 TSSNLKIHARKHSGEKPYKCDRCQGRFAQA---STLYHVRALLBEKPYVCTCGKA 452
 QY 428 LKSPSPQVAHLRAHVLIIHTGEKPYCPTGTRFRHLITLTKSHVRIHTGEKPYHCDCPGLH 487
 Db 453 FAVSSSLITHSKR---HTGEKPYICGICGKSFSSGELNKHFRSHTGTRPITCELGNS 508
 QY 488 FRHKSQRLHLRQKHGAATNT 508
 Db 509 YTDINKLKHKTKVHSGADKT 529

RESULT 13
 ABB50159
 ID ABB50159 standard; Protein: 610 AA.
 XX
 AC ABB50159;
 XX
 DT 05-FEB-2002 (first entry)
 DE Human transcription factor TRFX-10.
 XX
 XX Human; transcription factor; TRFX; cell proliferative disease;
 KM autoimmune disease; inflammation; neurological disease;
 KM developmental disorder; cancer; AIDS; infection; cytostatic; anti-HIV,
 KM neuroprotective; antiinflammatory; gene therapy.
 OS Homo sapiens.
 PN WO200172777-A2.
 PD 04-OCT-2001.
 PF 13-MAR-2001; 2001WO-US08117.
 PR 13-MAR-2000; 2000US-0188986.
 XX


```

Db      2 QVSHHCEHLERLNKQREAGFLCDCTIVIGEFQFAHRNVLASFSEYFGAIYRSTSENNV 61
      74 GVDVLSLPGGPEARGFAPLIDPMYTSRLRLSPATPAVLAATYLOMEHVQACH----- 128
      62 FLDSQOV---KADGFOKLEPIYTGTLNDSMNWKEIHQAADYLVKEEVVTKCKIKMED 117
      129 -----RFIOASYEPLGISLRP-----LEAPPTPTAPPP 159
      118 PAFIANPSSTEISITGTELNQOTCLTLRDYNNREKSEVSTDLIQANPKGALAKKSS 177
      160 -SPRSEGHDPPTRESRSCSQP-----PSPASPPKACNMK---KY 197
      178 QTKKKKKAFFNSPKTGQNKTVQYPSDILENASVELFLDANKLPTPVVEQVAQINDNSELEL 237
      198 KYIVLNSQASQ--AGSLVGERSSGQPCPOARLPS---GDEASSSSSSSSSEEGPIPG 252
      238 TSVENTFPADIVHTVTVKRGKSGOPNCALKEHSMNSNIASVKSPLYEANSGBE--LDQ 295
      253 POSRLSPTATVQFCGAPASTPYLLTSQAODTSGSPSERARP---LPSEFFSCQNC 308
      296 RYSKAKPMCNT---CGKVFSEASSLRHRMR.IHKG-----VKPYCHLGGKAFQCNQOLK 346
      309 AVAGCSSGLDSLVPGEDDKPYKCOLCRSFRYKGNLASHRTVHTG-EKPYHCSICGARP 367
      347 THVRTHTG-----EKPYKCELCDFKQFACQCVLFHSGRMHGEKPYKCDVNLQGA 397
      368 RPAHLKTHSRINSGEKPYKCECTGSRFVQVRSQPPSGFGKPARAGVGQKGFCCSSQRD 427
      398 TSSNLKIHKRKHSGEKPYVCDRCGQRF-----424
      428 LKSPSQVLAHLRAVLIHTGEKPYPCPTGTRFRHLQTLKSHVRIHTGEKPYHCDPCGLH 487
      425 ---AQASTLYTVHVRHTGEKPYVCDTCGKAFVASSSLITHSRKHTEKPYICGICGKS 479
      488 FRHKSQRLHLRQKHG 503
      480 FISSGELNKHFRSHTG 495

RESULT 15
ID AAM47779 standard; Protein; 610 AA.
AC AAM47779;
DT 01-MAR-2002 (first entry)
DE Human myoneurine.
KW Human; myoneurine; cytosolic; neurological; muscular; gene therapy;
KW cancer; neuromuscular disease; dystrophy; polymyositis; rhabdomyosarcoma;
KW BTB/POZ domain; zinc finger domain.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FT Domain 1..111
XX FT /note= "BTB/POZ domain. This sequence is also given
XX FT in AAM47780"
XX FT 304..527
XX FT /note= "zinc finger domain. This sequence is also given
XX FT in AAM47781"
XX
XX PN FR2808535-A1.
XX
XX PD 09-NOV-2001.
XX
XX PF 05-MAY-2000; 2000FR-0005792.
XX
XX PR 05-MAY-2000; 2000FR-0005792.
XX
XX PA (ALILI) ALLTEL P M.

```

```

XX Seddiq N, Cluences DC, Rieger F, Perin JP;
PI
XX MPI: 2002-019508/03.
XX DR N-PSDB; ABA04825.
XX
XX PT New nucleic acid encoding human myoneurin, useful for diagnosis,
XX PT prognosis and treatment of cancer and neurological diseases, comprises
XX PT a polynucleotide encoding the human myoneurin protein -
XX
XX PS claim 18; Fig 2; 64pp; French.
XX
XX CC The present sequence is the protein sequence for human myoneurin. The
XX CC myoneurin coding sequence and its fragments are useful as primers and
XX CC probes, for diagnosis, prognosis or assessing risk of human diseases,
XX CC particularly cancers and neuromuscular diseases (e.g. dystrophy,
XX CC polymyositis or rhabdomyosarcoma). The coding sequence and its fragments
XX CC are also useful in gene therapy for these diseases.
XX
XX SQ Sequence 610 AA;
XX
XX Query Match 17.3%; Score 482; DB 23; Length 610;
XX Best Local Similarity 26.8%; Pred. No. 7.3e-26;
XX Matches 149; Conservative 65; Mismatches 212; Indels 130; Gaps 16;
XX
XX 16 EFTRHSSDVLGNINELRLGLTDTVTLVGGQPLRAHKAVALIACGFFYSIRGRA--GV 73.
XX 2 QVSHHCEHLERLNKQREAGFLCDCTIVIGEFQFAHRNVLASFSEYFGAIYRSTSENNV 61
XX
XX 74 GVDVLSLPGGPEARGFAPLIDPMYTSRLRLSPATPAVLAATYLOMEHVQACH----- 128
XX 62 FLDSQOV---KADGFOKLEPIYTGTLNDSMNWKEIHQAADYLVKEEVVTKCKIKMED 117
XX 129 -----RFIOASYEPLGISLRP-----LEAPPTPTAPPP 159
XX 118 PAFIANPSSTEISITGTELNQOTCLTLRDYNNREKSEVSTDLIQANPKGALAKKSS 177
XX 160 -SPRSEGHDPPTRESRSCSQP-----PSPASPPKACNMK---KY 197
XX 178 QTKKKKKAFFNSPKTGQNKTVQYPSDILENASVELFLDANKLPTPVVEQVAQINDNSELEL 237
XX 198 KYIVLNSQASQ--AGSLVGERSSGQPCPOARLPS---GDEASSSSSSSSSEEGPIPG 252
XX 238 TSVENTFPADIVHTVTVKRGKSGOPNCALKEHSMNSNIASVKSPLYEANSGBE--LDQ 295
XX 253 POSRLSPTATVQFCGAPASTPYLLTSQAODTSGSPSERARP---LPSEFFSCQNC 308
XX 296 RYSKAKPMCNT---CGKVFSEASSLRHRMR.IHKG-----VKPYCHLGGKAFQCNQOLK 346
XX 309 AVAGCSSGLDSLVPGEDDKPYKCOLCRSFRYKGNLASHRTVHTG-EKPYHCSICGARP 367
XX 347 THVRTHTG-----EKPYKCELCDFKQFACQCVLFHSGRMHGEKPYKCDVNLQGA 397
XX 368 RPAHLKTHSRINSGEKPYKCECTGSRFVQVRSQPPSGFGKPARAGVGQKGFCCSSQRD 427
XX 398 TSSNLKIHKRKHSGEKPYVCDRCGQRF-----424
XX 428 LKSPSQVLAHLRAVLIHTGEKPYPCPTGTRFRHLQTLKSHVRIHTGEKPYHCDPCGLH 487
XX 425 ---AQASTLYTVHVRHTGEKPYVCDTCGKAFVASSSLITHSRKHTEKPYICGICGKS 479
XX 488 FRHKSQRLHLRQKHG 503
XX 480 FISSGELNKHFRSHTG 495

```

Search completed: July 14, 2003, 18:14:58
Job time : 21.9594 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:11:27 ; Search time 5.81542 Seconds

(without alignments)
2620.804 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789
Sequence: 1 MGSPAPGALGVREFTRH.....ROKGAATVTKVHILGSP 518Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents, AA:*
1: /cgn2_6/prodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/1/1aa/Backfile1.fep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1095.5	39.3	706	1	US-08-074-967-2
2	1095.5	39.3	706	2	US-08-553-541B-2
3	1095.5	39.3	706	4	US-09-268-202-2
4	1095.5	39.3	706	5	PCT-US94-06669-2
5	576.5	20.7	803	4	US-09-063-035-2
6	419.5	15.0	488	2	US-08-933-750C-17
7	419.5	15.0	488	4	US-09-234-613-17
8	384.5	13.8	547	1	US-08-340-203A-3
9	384.5	13.8	547	2	US-08-452-567-3
10	384.5	13.8	547	2	US-08-452-427-3
11	384.5	13.8	547	3	US-09-085-407-3
12	375.5	13.5	671	3	US-09-121-321-16
13	375.5	13.5	671	4	US-08-933-803A-16
14	362	13.0	543	4	US-09-362-123A-4
15	357.5	12.8	711	2	US-08-820-170A-10
16	357.5	12.8	711	4	US-09-055-699-10
17	357.5	12.8	711	4	US-09-273-565-10
18	357.5	12.8	711	4	US-09-565-538-10
19	357.5	12.8	711	4	US-09-561-468-10
20	348.5	12.5	462	3	US-08-486-099-117
21	348.5	12.5	462	3	US-08-360-107A-117
22	348.5	12.5	462	3	US-08-484-223B-117
23	348.5	12.5	462	3	US-08-919-597-117
24	348.5	12.5	462	3	US-08-475-668A-117
25	348.5	12.5	462	3	US-08-485-551A-117
26	348.5	12.5	462	3	US-08-471-913A-117
27	348.5	12.5	462	4	US-08-485-264A-117

28	348.5	12.5	462	4	US-08-474-349A-117	Sequence 117, App
29	339	12.2	181	4	US-08-676-318A-44	Sequence 44, App1
30	329	11.8	181	4	US-08-863-813A-44	Sequence 44, App1
31	305	10.9	727	2	US-08-475-844-9	Sequence 9, App1
32	305	10.9	727	5	PCT-US95-08429-9	Sequence 9, App1
33	305	10.9	728	2	US-08-475-844-5	Sequence 5, App1
34	305	10.9	728	5	PCT-US95-08429-5	Sequence 5, App1
35	302	10.8	112	6	5206152-4	Patent No. 5206152
36	299	10.7	338	2	US-08-933-750C-4	Sequence 4, App1
37	299	10.7	338	4	US-09-234-613-4	Sequence 4, App1
38	296.5	10.6	648	4	US-09-262-773-2	Sequence 2, App1
39	296.5	10.6	648	4	US-09-262-773-2	Sequence 2, App1
40	289	10.4	110	2	US-08-553-541B-9	Sequence 9, App1
41	289	10.4	110	4	US-09-268-202-9	Sequence 9, App1
42	281	10.1	106	1	US-08-340-203A-10	Sequence 10, App1
43	281	10.1	106	2	US-08-452-427-10	Sequence 10, App1
44	281	10.1	106	3	US-09-085-407-10	Sequence 10, App1
45	266.5	9.6	429	1	US-08-234-783-4	Sequence 4, App1

ALIGNMENTS

RESULT 1

US-08-074-967-2
; Sequence 2, Application US/08074967
; Patent No. 5641672

GENERAL INFORMATION:

APPLICANT: Dalla-Favera, Riccardo
APPLICANT: Chaganti, R.S.K.
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10112

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/074,967
FILING DATE:
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 0575/43771
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525

TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 706 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
US-08-074-967-2

Query Match 39.3%; Score 1095.5; DB 1; Length 706;
Best Local Similarity 35.7%; Pred. No. 3.4e-76;
Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;

QY 1 MGSPAPGALGVREFTRHSDVGNLRLRGLTDVTLVGGQPLRAKAVLACS 60
DB 1 MASPA---DSCI---QFTRHADVLLNLRLSRDLTDVIVVSRQFRAKTVLQMR 54

Best Local Similarity 35.7%; Pred. No. 3:4e-76;
Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18

QY	MSPPARBEALGVYREFTSHSSVDLGNLWELRLRGILTTDTVLVGGOPJBAHXAVYL IACS	60
Db	1 MASPA--DBCT-----OPTHARDVLNLRNLSRDLITDVIYVSSBOQFAHHTVIMAMR	54
QY	61 GFFYSIFRGAGVGVLDVLSLPGGPEARGAFLPDDFWYTSRLRLSPATAVLAATYLOM	120
Db	55 GLFYISFTDQLKCNLSVIMLDEINEGCIIIDFWYTSRLNLRBGNIMAVMATYLOM	114
QY	121 EHVVAQCHRFIOAS-----	134
Db	115 EHVVDTCRKFIXASAEAMVSAIKPPREBFANSHMLPMODIMAYRGREVENNLPLRSAPG	174
QY	135 -----YEBLGIS-----	141
Db	175 CESRAFAFSLYGLSTIPPASYSMTSHLPVSSLFSDBEFFRDVMPVANNPPKERALPCDS	234
QY	142 -----LRPL-----	145
Db	235 ARPVGEVGRPTLLEVS PNVCHSNISYPKETIPEARSDMHYSVABGLKAPAPARNAAYE	294
QY	146 -----EAEPTT-----EPAP-----DPGSPRSEGHDPDPTES-----RSC	177
Db	295 PCDKASKEBERSESSEDEIALHFEPPAPAPLNKGLVSPQSFOKSDCQPNSEPTFACSSKNAC	354
QY	178 -----GQSPSPASPDPOKACNMKKYKIVYNS-----QASQASLJVGERSSQPC	223
Db	355 IIOGSGSPPAKSPITDPAKCSMKYKIVYNSLNLQNAKPGGBOALEGRISPRAYTAPAC	414
QY	224 QARLPSEGADEASSSSSSSSSSEEGPIPGOSRL-----SEPTAATVCP-----	266
Db	415 QPPM--EPENLQDSPTKLSASGDSITP--QASLNNIIVKMSMTGSPRSSSESHSLYMPH	472
QY	267 ---KCG--APASTPYLLTSQ---AOTSQSESEPARPLPGSEFFSPCONCEAVACSSG	316
Db	473 PKCTSCGSGSPHAEWCLHTAGPTFAEMGETOSBYSDSSCEGAFFCNECDRFSBEAS	532
QY	317 LD-SLVPGDEDPYKCOLCRSSSPRYGNLASHRVTYTGKEPYCNISCGARFNPNALUKYH	375
Db	533 LGRHLLQTHSDRPYKCDRCOASFRYVGNLASHTVTYTGKEPYFPCNITCGAOFNPNALUKYH	522
QY	376 SRHISGEKPYKCECTCGSRFVOVRSQPPSGFOGKPARGVGOUKGAFCSQRODLKSPSQY	435
Db	593 TRHISGEKPYKCECTGARFV-----	614
QY	436 AHLRAHVLITHGEKPYPCPTCGTRFPHLOTLSHVRIITHGEKPYHCDPCGLHFRHKSQLR	495
Db	615 AHLRAHVLITHGEKPYPCPCICGTRFPHLOTLSKHLRIITHGEKPYHCEKCNLHFRHKSQLR	674
QY	496 LHLROKGAATNTKYVYHI 514	
Db	675 LHLROKGAATNTKYOYRV 693	

RESULT 5
US-09-063-035-2
; Sequence 2, Application US/09063035
; Patent No. 6150001

```
?
? GENERAL INFORMATION:
? APPLICANT: PEUKERT, Karen; HANEL, Frank; and EILERS,
? APPLICANT: Martin
? TITLE OF INVENTION: Myc-binding zinc finger proteins
? TITLE OF INVENTION: their preparation and their use
? NUMBER OF SEQUENCES: 2
? CORRESPONDENCE ADDRESS:
?
```

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.2 mb storage
COMPUTER: IBM AT-compatible, 80486 processor
OPERATING SYSTEM: MS-DOS version 6.1

CURRENT APPLICATION DATA:

CLASSIFICATION: 514
INFORMATION FOR SEQ ID NO: 2

LENGTH: 803 amino acids

US-09-063-035-2

Query Match	20.7%;	Score 576.5;	DB 4;	Length 803;
.Best Local Similarity	29.4%;	Pred. No. 2.8e-36;		
Matches 180;	Conservative 64;	Mismatches 205;	Indels 163;	Gaps 20

Qy	16	EFRRHSADVLMNINLELIGLITVDLTVLLVGCQPLRANHKAVALIACSGFFYIIFRRAG-VG	74
Dd	2	DFQHSQNVLEQNLQCRQLGLCDCTFVVDGVNHRKAKHVAVLAASVEYFKYLFVDQKDVH	61
Qy	75	VDVLSLPGREARGFARLDEMYTSTRLSLSPATAPAVLAATYVLOMEHVVOACHRFIOAS	134
Dd	62	LDI-----SNAAGLGQMLEFMFYTAKLSTLSEENVDVLAATVFLQMODITYACH-----	109
Qy	135	YEBLGISLRPLAEPRTP-----TAPPSPPRSSECHPRPRPESSSCQSPSPASPD	188
Dd	110	-----ALXSL-ABEPTSPGNAELATEGGDKAKEBKATSTLSRLBEQAGSTPLGP-161	
Qy	189	PKACNNKKYKTYIVLNSOASQAGSLVGRSSGQPCPOARLP-----SGDEASSSSSSSSSSSS	245
Dd	162	-----SRDLKERGGQQAASASGAEQTEKADAPREPRVELKDPRTSGMAAABEABALSSSE	219
Qy	246	EEBPT-PEGQSLSPATVTPQKCA-PAST-----PYLITSQA-----QDTGSPF--	220
Dd	220	QEMEVVPARKGEEQKEQEBQEBEAGAEVKEGSOLENGAEPEENEENESAGTSGQE	279
Qy	291	---BRAPLP-----GSSF-----FSCONC	307
Dd	280	LGSEANGLASGYGDRTESKAVGSYHNKCEQOGELFHTNGFKHRIHNHGEKFSCREC	339
Qy	308	EAV-----ACSSGLDSLVPGEDEKPKYQLCRSSFRY-----	340
Dd	340	SKAFSPDAPACKAHEKTHSP-----LKPYGCEGCKSGKSYRLISLNLHKRHSGEARYRCEDCG	366
Qy	341	-----KGNLASRPTYHTEKPYHNGICG-----AREN	367
Dd	397	KLFETTSGNLKRQVLVHSGEKPYQCDVCGRSFSDPTSKRMLETHDTEKHKCPHCDKFN	456
Qy	368	RPANLKTGSRHSIGGEKPYCETGCSGFQVNSOPRPSGFGCKPARGVGQKGFSSQROD	427
Dd	457	QVGNLKAHLKIHLADGRLKCRGCGQFL-----TSGNLKQQLRHSIGGEKPYCINHOQRO	510
Qy	428	LKSPBSQVAAHRAVULIHTGEKPYCPTGCGRPHNLQTLKSHVRIHTGEKPYHDCPSGLH	487
Dd	511	FADPGA-----LQRHRIHTGEKPCQCVWACGKAFTQASLSLHVQHTHGEKPYCERCGKR	566
Qy	488	PHKSGQLRLHLR	499
Dd	567	FVQSSQLAHNHR	578

RESULT 6
US-08-933-750C-17
; Sequence 17, Application US/08933750C
Date: 08/08/2008 10:00:00

GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L
APPLICANT: Bandman, Olga

APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750C
FILING DATE: September 23, 1997
CLASSIFICATION: 536
PRIORITY APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 488 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRISTNOT03
CLONE: 641127
US-08-933-750C-17

Query Match 15.0%; Score 419.5; DB 2; Length 488;
Best Local Similarity 29.5%; Pred. No. 1.7e-24;
Matches 110; Conservative 36; Mismatches 132; Indels 95; Gaps 12;

176 SCSQPPSPASPDPKACN-----WKYKYIYVLSOA-----SQAGSLVGRSS 218
137 TCQNG-QLKESLDPIDCKNDKIDHWK--SQVSCSQQRGHTBEKPCDHNNCKIILNTSPD 193
219 GPCPCQARLPSCGDEASSSSSSSSSESGPIPGQSRUSPTAAVQFKGAPASTPYLL 278
194 GHPYEKIHAEKQYEGSCGCKNFSQSS-----LL 223
279 TSOAODTSSPSERRARPLPGSEFFSCQNC-EAVAGCSSGLDVLPGDDEKPYKCOLCRSS 337
224 LHQRDHTKEP-----YKCEQCGKGFTRSSSLIHQAVHTDEKPYKCDKCGKG 271
338 FRKYNLASHRTVHTGEKPYHCSICGARFNRPAHLKTHSHIGSEKPYKCEGSRFVQ- 396
272 FTRSSSLILHNAVHTGEKPYKCDKCGKFSQSKLIHQVHTGEKPYCEGEGMFSQR 331
397 ---VRSQPSGFGKPARGVGQKCGFSSSQRODLKSPSQVAHLRAVLIHTGEKPY 452
332 SNLHIHQVHTG--ERRPYKCGECGKF-----SQSSNLHIHRCIHTGEKPYQ 376
453 CPTGGRFPHLQTLKSHVRIHTGEKPYHCDPQGLHFRHKSQ-----RLHLRK- 501
377 CTECGGFSQSSDLRLHLRVHTGEKPYHCGKCGKFSQSSKLIHQVHTGEKPYKCDKCGK 436

QY 502 -HGAANTKRVHYH 513
DB 437 GKGFSQSSNLHIH 449

RESULT 7

US-09-234-613-17
Sequence 17, Application US/09234613
Patent No. 6132973

GENERAL INFORMATION:

APPLICANT: Lal, Preeti
APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Shah, Purvi
APPLICANT: Au-Young, Janice
APPLICANT: Yue, Henry
APPLICANT: Guegler, Karl J.
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/234,613
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/933,750
FILING DATE: September 23, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 488 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRISTNOT03
CLONE: 641127
US-09-234-613-17

Query Match 15.0%; Score 419.5; DB 4; Length 488;
Best Local Similarity 29.5%; Pred. No. 1.7e-24;
Matches 110; Conservative 36; Mismatches 132; Indels 95; Gaps 12;

176 SCSQPPSPASPDPKACN-----WKYKYIYVLSOA-----SQAGSLVGRSS 218
137 TCQNG-QLKESLDPIDCKNDKIDHWK--SQVSCSQQRGHTBEKPCDHNNCKIILNTSPD 193
219 GPCPCQARLPSCGDEASSSSSSSSSESGPIPGQSRUSPTAAVQFKGAPASTPYLL 278
194 GHPYEKIHAEKQYEGSCGCKNFSQSS-----LL 223
279 TSOAODTSSPSERRARPLPGSEFFSCQNC-EAVAGCSSGLDVLPGDDEKPYKCOLCRSS 337
224 LHQRDHTKEP-----YKCEQCGKGFTRSSSLIHQAVHTDEKPYKCDKCGKG 271

Qy	338	FRKYNJASRVTHTGEKPYHSCICAPRNPANIKTSRHSIGCEPKYCETCSRPVQ-	3366
Db	272	FTYSSSLIHNAVHTGEKPYKCDKCGKGSQSKLTHORVHTGEKPYCECGCMSSOR	3311
Qy	397	---VBSQPSGFGQKPARGVGQKGFSSRQDLKSPSQVAHLRAVLIHTGEKPY	452
Db	332	SNLIHQRVHTG--ERPYKCGEGCGKF-----SQSSNLHHRCHTHTGEKPYQ	3766
Qy	453	CPFGCGRFRMLQTLKSHVRIHTGEKPYHCDPQGLHFRKXSQ----	501
Db	377	CYECGKGFSSQSDLRIHLRVHTGEKPYHCGKCGKGSQSKLTHORVHTGEKPYECSKC	436
Qy	502	-HGAATNTKVHYH	513
Db	437	GKGFSSQSNLHITH	449

```

US-08-340-203A-3
Sequence 3 Application US/08340203A
Patent No. 5756668
GENERAL INFORMATION:
APPLICANT: Baylin, Stephen B.
APPLICANT: Wales, Michele M.
TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: California
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/340,203A
FILING DATE: 15-NOV-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.
REGISTRATION NUMBER: P-38,347
REFERENCE/DOCKET NUMBER: 07265/039001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 455-5100
TELEFAX: (619) 455-5110
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 547 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-340-203A-3

```

Query Match	13.8%;	Score 384.5;	DB 1;	Length 547;
Best Local Similarity	24.6%;	Pred. No. 9.4e-22;		
Matches 128;	Conservative 54;	Mismatches 170;	Indels 169;	Gaps 17;

```

QY 20 HSSDVLGNLNEBLRGLTLDVTLVGGQLRAKAVLLACGGFYSLFRGAGVGVNLS 79
Db 10 HSRQLLQINNOTKFLCDVIIWQNALFRHKVLLAASAYLKLSLV----VHDLIN 64
QY 80 LPGPEARGLFAFLDPMYTSRL-----RLSPATP---AVLAATYLOMEHYVQ 125
Db 65 LDDHWSPAVFLVLDFITYGLRGLADGAERAAAAAANAVPAEBSLGAIVLAALAASTYQIDPLVA 124
QY 126 ACRHFTQ-----ASYEP-----LGISLRLEBAEPPTTPAEP-PCGSPR 162
Db 125 LCKKRLKRRHKYCHLRGGGGGGGGYAPAMATSWAGSAAPASAKSVGTGPRPCGPR 184

```

```

QY 163 RS-----EGHDDPTESSCSGSPSPSPADPACMKWKYKTVILNSQASQA 209
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 185 SANRRRATLLAMWTGPAPARATATTTTAAARPPVAARPPAP----- 225
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 210 GSLVGERSSGCPQPOAR-----LPQGDENASSSSSSSS--SSSSEGPJGP 253
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 226 ----GGHLEGYICPHIATVGPESPFDNLVYCI PCGKGPPSEBQJNAHVAHVE----- 275
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 254 QSRHSPTATVQFKGAPASTPYLLTQAODTSGSPSERARPLPGESEFQCNCEAVAGC 313
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 276 BEALYGRAEAALVAAGAAAGLGP-----PFGGGG-----DKVAGA 309
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 314 SSGSLDVLVEGDEDEKPRKQOLCRSSPRYKGNLASHRTYHTGSEPRHCSICGARFVRPNLK 373
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 310 PGGGLGELL-----RPYRCGSCDKSKSYKDPATLROHEKTHMLTRPYPCITCGKFTQRGTM 364
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 374 TSGRLHSGEKPYKCEKCSRFVQVNSQPPSPGQCKPARGVGQKGFCSQRODLKSPS 433
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 365 RHRRSHLGLKPPACDAGCGRRPTR----- 387
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 434 QVAMLRATVLIH-TGEKPYPCPTCGTRERHLOTLKSHVRIH 473
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 388 QYRLTRTHMRIRHPGEKPEYECQVCGSKRAQQRANLISHKQH 428
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

1      RESULT 9
2      US-08-452-567-3
3      Sequence 3, Application US/08452567
4      Patent No. 5846712
5      GENERAL INFORMATION:
6      APPLICANT: Baylin, Stephen B.
7      APPLICANT: Males, Michele M.
8      TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
9      NUMBER OF SEQUENCES: 3
10     CORRESPONDENCE ADDRESS:
11     ADDRESSEE: Fish & Richardson P.C.
12     STREET: 4225 Executive Square, Suite 1400
13     CITY: La Jolla
14     STATE: California
15     COUNTRY: USA
16     ZIP: 92037
17     COMPUTER READABLE FORM:
18     MEDIUM TYPE: FLOPPY disk
19     COMPUTER: IBM PC compatible
20     OPERATING SYSTEM: PC-DOS/MS-DOS
21     SOFTWARE: PatentIn Release #1.0, Version #1.25
22     CURRENT APPLICATION DATA:
23     APPLICATION NUMBER: US/08/452,567
24     FILING DATE: 25-MAY-1995
25     CLASSIFICATION: 435
26     ATTORNEY/AGENT INFORMATION:
27     NAME: Halle, Ph.D., Lisa A.
28     REGISTRATION NUMBER: 38,347
29     REFERENCE/DOCKET NUMBER: 07265/039003 (DIY PD3664)
30     TELECOMMUNICATION INFORMATION:
31     TELEPHONE: (619) 678-5070
32     TELEFAX: (619) 678-5099
33     INFORMATION FOR SEQ. ID NO: 3:
34     SEQUENCE CHARACTERISTICS:
35     LENGTH: 547 amino acids
36     TYPE: amino acid
37     TOPOLOGY: linear
38     MOLECULE TYPE: protein
39     US-08-452-567-3

```

Query Match 13.8%; Score 384.5; DB 2; Length 547;
Best Local Similarity 24.6%; Pred. No. 9,4-22;
Matches 128; Conservative 54; Mismatches 170; Indels 159; Gaps 17

OY 20 HSDVLCNLNELRLGLTDTVLIVGGPFLRAHVALIACSGFYSTRGRAGVCVDYLVS 79
::: :::
10 HSRIQLQNNORRKGKGLCDVIIYNQALFFAHKNVLAASAYAKSLV-----YHDNLLIN 64
::: :::

```

QY      80 LFGPEARG-FAPLIDFWYTSRL-----RUSPATAP---AVLAATYLQNEHVQ 125
      65 LDHDMVSPAVFVLVDIFYTGRLDGABAAAAAVALPFAEPBISGAVLAASVYLQIPDLVA 124
QY      126 ACHRIQ-----ASYEP-----LGISLRLPEAEPPPTAPP--PGSR 162
      125 LCKKRLKRGKXCHLRRGGGGGGGYAPYAMATSWAGSAAPPAKSVVGTSPSRPGPR 184
QY      153 RS-----EGHPDPPTESRSCSQGPPSPASPDPKACNMKKYKYLINSQASQA 209
      155 SAMRRRRAATLAAMTGPAPAAATTTTAAARRPVAARTAP--P----- 225
QY      210 GSLVERSSGQPCQAR-----LPSGDEASSSSSSS--SSSEEGPIPGP 253
      226 ----CGHLEGYPCPHLAYGEPESFGDNLVYCIPCGKGFPSBQLNAHVAHYE----- 275
QY      254 QSRLEPTAATVQFKGAPASTPYLLTQAODTSSGSEBARPLPSSEFPSCONCEAVAC 313
      276 BEALYGRAEAAVAAGAGLGP-----PFGGGG-----DKVAGA 309
QY      314 SSGLDLVPGDEBDEKPYKCOLCRSSPRYKGNLASHTVHTGKPYHCSICGAFNRPANLK 373
      310 PGGLGELL-----RPRCGSCDKSKYKDPATLQHEKTHMLTRPYCTICGKFTQRGIT 364
QY      374 THSRHSGEKPYKCTCGSRFVQVRSQPPSGFGKPARGVGQKGFSSORQDLKSPS 433
      365 RMRSHLGLKPPACACGMRFT----- 387
QY      434 QVAHLRAVLIH-TGEKPYPCPTCGTRPRHLQTLKSHYRIH 473
      388 QYRLTRTHRIHPREKEPYECVCGKFAQORNLISHMKH 428
Db

```

RESULT 10

```

US-08-452-427-3
; Sequence 3, Application US/08452427
; Patent No. 5922590
; GENERAL INFORMATION:
; APPLICANT: Baylin, Stephen B.
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Releasee #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,427
; FILING DATE: 26-MAY-1995
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/340,203
; FILING DATE: 15-NOV-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Ph.D., Lisa A.
; REGISTRATION NUMBER: P-38,347
; REFERENCE/DOCKET NUMBER: 07265/039001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 455-5100
; TELEFAX: (619) 455-5100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 547 amino acids
; TYPE: amino acid

```

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-427-3

Query Match 13.8%; Score 384.5; DB 2; Length 547;

Best Local Similarity 24.6%; Pred. No. 9,4e-22;

Matches 128; Conservative 54; Mismatches 170; Indels 169; Gaps 17;

```

QY      20 HSDVILGNINLEIRLGITLDVTLVGGQPLBAHKAVALIACSGFFSIFRGRAGVGVYLS 79
      10 HSRQLDLQNNQRTGFLCDVIIVQNALFRHKRVLAASSAYLKSIV-----VNDNLIN 64
Db
QY      80 LFGPEARG-FAPLIDFWYTSRL-----RUSPATAP---AVLAATYLQNEHVQ 125
      65 LDHDMVSPAVFVLVDIFYTGRLDGABAAAAAVALPFAEPBISGAVLAASVYLQIPDLVA 124
QY      126 ACHRIQ-----ASYEP-----LGISLRLPEAEPPPTAPP--PGSR 162
      125 LCKKRLKRGKXCHLRRGGGGGGGYAPYAMATSWAGSAAPPAKSVVGTSPSRPGPR 184
QY      153 RS-----EGHPDPPTESRSCSQGPPSPASPDPKACNMKKYKYLINSQASQA 209
      155 SAMRRRRAATLAAMTGPAPAAATTTTAAARRPVAARTAP--P----- 225
QY      210 GSLVERSSGQPCQAR-----LPSGDEASSSSSSS--SSSEEGPIPGP 253
      226 ----CGHLEGYPCPHLAYGEPESFGDNLVYCIPCGKGFPSBQLNAHVAHYE----- 275
QY      254 QSRLEPTAATVQFKGAPASTPYLLTQAODTSSGSEBARPLPSSEFPSCONCEAVAC 313
      276 BEALYGRAEAAVAAGAGLGP-----PFGGGG-----DKVAGA 309
QY      314 SSGLDLVPGDEBDEKPYKCOLCRSSPRYKGNLASHTVHTGKPYHCSICGAFNRPANLK 373
      310 PGGLGELL-----RPRCGSCDKSKYKDPATLQHEKTHMLTRPYCTICGKFTQRGIT 364
QY      374 THSRHSGEKPYKCTCGSRFVQVRSQPPSGFGKPARGVGQKGFSSORQDLKSPS 433
      365 RMRSHLGLKPPACACGMRFT----- 387
QY      434 QVAHLRAVLIH-TGEKPYPCPTCGTRPRHLQTLKSHYRIH 473
      388 QYRLTRTHRIHPREKEPYECVCGKFAQORNLISHMKH 428
Db

```

RESULT 11

```

US-09-085-407-3
; Sequence 3, Application US/09085407
; Patent No. 6103877
; GENERAL INFORMATION:
; APPLICANT: Baylin, Stephen B.
; TITLE OF INVENTION: NOVEL TUMOR SUPPRESSOR GENE, HIC-1
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: California
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Releasee #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/085,407
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,203
; FILING DATE: 15-NOV-1994

```

```

?      ATTORNEY/AGENT INFORMATION:
?      NAME: Haile, Ph.D., Lisa A.
?      REGISTRATION NUMBER: P-38,347
?      REFERENCE/DOCKET NUMBER: 07265/039000
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: (619) 455-5100
?      TELEFAX: (619) 455-5110
?      INFORMATION FOR SEQ ID NO: 3:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 547 amino acids
?      TYPE: amino acid
?      TOPOLOGY: linear
?      MOLECULE TYPE: protein
US-03-085-407-3

```

Query Match	13.8%	Score	384.5	DB	3	Length	547
Best Local Similarity	24.6%	Pred	No.9.4e-22				
Matches	128	Conservative	54	Mismatches	170	Indels	169
						Gaps	177

```

Qy 20 HSSSVLGNLNELELRGLDPTLVLLVGGCPLRAHKVLLACSGFFFSIFRGAGVGVLS 79
Db 10 HSNQQLLQLNQNTKGLCDVIIIVGNALFPAHKVLAASSAYLKSIV-----VHDMLN 64

Qy 80 LPGGPEARG-FAPLDPMYTSRI-----RUSPATAP--AVTAAATYLOMEHVQ 125
Db 65 LDHDMVSPAVFRLVDFITYGRLADGAEEAAAAAAVAPGAESPLGAVLAAAYLOIDPLVA 124

Qy 126 ACHRFIO-----ASYEP-----LGISLRPLEAEPPTPAAP-PQSPR 162
Db 125 LCKKRLKRNKRYCHLAGGGGGGGAIFYAMATSMGSAAPPASNAKSVYGRBPRPGPR 184

Qy 163 RS-----EGHDDPTESHSCSGGPPSPADPKACMWKYIVLNSQASQA 209
Db 185 SAMRRRATLAATGPARAATAATTTAAARRRVAARTAPP-----225

Qy 210 GSVLGERSSGQPPQAR-----LPQDDEASSSSSSSS--SSSSEGP1RG 253
Db 226 ----GGHLEGPCHPLAYGPESFGNLVYC1PCGKGPPSSQULNAHYEAIVE-----275

Qy 254 QSRLSPTAATVQPKCAPASTPYLLTSAQDPTSGSPSERARLPQSEFFSCONCEAVAGC 313
Db 276 EEAALYGRAAEAEEVAAGAAGLGP-----PFGGCG-----DKVAGA 309

Qy 314 SSGGLDLYPDEDDEKPYKCOLCRSSPFYKKNLSHRTVHTGEPKYCS1CGAREMRANLK 373
Db 310 PGLIGELL-----RPRCGSCDKSYDPAETLQHKHTMLTRPPTCT1CGKKT1QRGTM 364

Qy 374 THSRHSGEKPXYCETCGSRFVQVRSQPSGCGKPARGGVQKGGFCSQRODLKSPS 433
Db 365 RHRMRSHGLKRPACDACGMRFTR-----387

Qy 434 QVALRAHVLIH--TGEKPYPCPTCGTRFRHLDTLKSRIH 473
Db 388 QYRLTRTHMR1PRGEKPYPCOVCGGKFAQGNLLSHMKH 428

```

RESULT 12-321-16
 US-09-121-321-16
 : Sequence 16, Application US/09121321
 : Patent No. 6090783
 : GENERAL INFORMATION:
 : APPLICANT: Saiga, Akihiko
 : APPLICANT: Orita, Satoshi
 : APPLICANT: Igataishi, Hisanaga
 : APPLICANT: Okumura, Kouichi
 : APPLICANT: Sakaguchi, Gaku
 : TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
 : TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
 : NUMBER OF SEQUENCES: 20
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: FISH & BEAVE
 : STREET: 1251 Avenue of the Americas
 : CITY: New York

STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/121,321
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/933,803
FILING DATE: 19-SEP-1997
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHGN-12CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9080
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 671 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-121-321-16

Query Match	13.5%	Score 375.5	DB 3	Length 671
Best Local Similarity	28.3%	Pred. No. 66-21		
Matches 106	26	Mismatches 114	Indels 129	Gaps 15

```

QY      15  PPAPRPGRSEGHDPPTESRSCSG-PP-----SPAPDPKACNMKKYKIVINSQ 205
Db      403  PPPAPRPQPPQKP-PPQPLQSQPPQSLPPPIAENPGSPBSG-----LLDD 455
QY      206  ASQAGSLVGRSSGQPCQAPRLPSGDEASSSSSSSSSEBPTPGQSRSLPTATVQ 265
Db      451  GFGV--LPBRSGE-AP-----PGDRLSTGGGGGGGGGG-----484
QY      266  FKCGAPASTPYLLTSQADTSSPSERRARPLPGSEFBSQNCENAVAGCSGL--DSLVPG 322
Db      485  ---GAAGCT-----GAGG-----GCCGC-----CPGGLRSLLLHG 512
QY      324  DEDKPYKQCLRRSPFYKGNLASHRTVTHGKPYHCISGARFMRPANKLTKHSRTHSGEK 383
Db      513  ANSKRPSCEBCCGKSGFVRKSLIYHRSHTKERYECACBCKSPFCHSGLLRHQNTHGKR 572
QY      384  PYSCETCSGRFYOVHSQPPSGFGQKPRAGVGQKGFCSQRODLKSPPSQVAHLRAVL 443
Db      573  PYSCSCEKTY-----SRKEHLQNHQ 594
QY      444  IHTGKPYCPCTGTFRHLQTLKSHVRIHTGKPYHCDPCGLHFRHKSQLRLTKQKG 503
Db      595  LHTGEPFPQALCGKSFIRKQNLKHORIHTGEPYTGEGECGKSFYKESLKHLLR---- 650
QY      504  AATNTKVYHILGPR 518
Db      651  -----VH---SGGP 656

```

RESULT 13
US-08-933-003A-16
Sequence 16, Application US/08933803A
Patent No. 6218522
GENERAL INFORMATION:
APPLICANT: Saiga, Akihiko
APPLICANT: Orita, Satoshi
APPLICANT: Igatahshi, Hisanaga
APPLICANT: Okumura, Kouichi
APPLICANT: Sakaguchi, Gaku

TITLE OF INVENTION: DNA MOLECULE RELATING TO SUPPRESSION OF
TITLE OF INVENTION: GENE EXPRESSION AND NOVEL PROTEIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: PISH & NEAVE
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/933,803A
FILING DATE: 19-SEP-1997
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Haley, James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: SHCN-12CIP
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 671 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-933-803A-16

Query Match 13.5%; Score 375.5; DB 4; Length 671;
Best Local Similarity 28.3%; Pred. No. 6e-21;
Matches 106; Conservative 26; Mismatches 114; Indels 129; Gaps 15;
QY 152 PPTAPPSPGPRSEGHDPPTESRSCSG-PP-----SPASPDPKACMKKYYIYLNQ 205
DB 403 PPPAPPPOPOPOPO-POPOLOSOPPOPSLPPIAAVENPGPPSRG-----LLDD 450
QY 206 ASQASLVGERSGQPCQPARLPSCGDEASSSSSSSSSSSEKPIPGQSRSLPTATVQ 265
DB 451 GRQV--LPERGSGE-AF-----PGGRSTGGGGGCGGGGG----- 484
QY 266 FPCGAPASTFYLLTQAQDTSGSPSERAPLPSEFFSCQCEAVAGCSSGL--DSLVPG 323
DB 485 ---GAEAGT-----GAGG-----GCGSC---CPGGLRRSLILHG 512
QY 324 DEDKPYKQCLSSPFYKGNLASHRTVHTGKPYHCSIGAPRNRPANLKTSHRSRHSR 363
DB 513 ASKPTSCPECGKSGFVRKSLIITHRSHTKRPYECABCEKSKNCHSLIRHOMTTRGR 572
QY 384 PYKCTCGSRFYQVNSQPPSGFGKPARGVGQKGFCSSDRDLKSPSOVAHLRAVYL 443
DB 573 PYKCSCEKTY-----SRKEHLQNHQR 594
QY 444 IHTGEPYPCPTGCTFRHLQTLKSHVRIHTGKPYHCDPGLHFNKSQLHLHQRKIG 503
DB 595 LHTGERPFCALCGKSPFRKQMLKQRIHTGERPYTCGECGKSPFYKESLMDHLR---- 650
QY 504 AATNTKVHYHILGSP 518
DB 651 -----VH---SGGP 656

RESULT 14
US-09-362-123A-4
Sequence 4, Application US/09362123A
Patent No. 6451558
GENERAL INFORMATION:
APPLICANT: Cooke, Michael Paul

APPLICANT: Holness, Claire
APPLICANT: Sirenk, Oksana
TITLE OF INVENTION: No. 6451558el Genes in the Control of Hematopoiesis
FILE REFERENCE: 4-30629A/SYS
CURRENT APPLICATION NUMBER: US/09/362,123A
PRIOR FILING DATE: 1999-07-27
PRIOR APPLICATION NUMBER: 09/128,310
NUMBER OF SEQ ID NOS: 7
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 543
TYPE: PRT
ORGANISM: Homo sapiens
US-09-362-123A-4

Query Match 13.0%; Score 362; DB 4; Length 543;
Best Local Similarity 27.1%; Pred. No. 4.9e-20;
Matches 127; Conservative 53; Mismatches 183; Indels 106; Gaps 18;
QY 102 RLSPATAVLAATYLMENHVQACH-----RFIOASYE--PLGISLRPLE----- 146
DB 86 RAFPDTPPGMTTS-----VFPVAGCHSVKSLQROKASPSREKRPVGSVYWERLLG 140
QY 147 -----AEPPTPAPPPSPRSE-----GHDPPTESRSCGPPS--PAS 186
DB 141 SGSGQASVLRITSPLRPEGVRLAEKTLTEHALGRQPRTERQKFCAGQVPGRTFGSA 200
QY 187 PDPKACN-----WKTKYIYLNQASQAGSL-----VGERS--SGQPCQARLPSG 230
DB 201 QDLAAGRGHHRMGAVQGEPRRLGQEPSTWDELGALHAGERSFCRCASKYFVSS 260
QY 231 DEASSSSSSSSSEEGPIPGQOSRLPTATVQFCKAPASTYLLTQAQDTSGSPS 290
DB 261 DLTKRLRHTG-----EPPECAQ-----CGKAFQSHLTQHRITSGEP 302
QY 291 BRARLPSESEPPSCQCEAVAGCSSGLDSLVPGE---DKPYKQLCRSFRYKGNLAS 346
DB 303 -----YACPVCGKAFRHS---SLVRHQRIHTAKSPRSCGKAPRSHSNLSQ 348
QY 347 HRTVTGKPYHCSIGAPRNRPANLKTSHRSRHSRHSRHSRHSRHSRHSRHSRHSR 406
DB 349 HRKIHAGRPVACQCGFRFCNLSHLIHERHTGKPYHCSIGAPRNRPANLKTSHRSR 403
QY 407 GKPARGVGQKGFCSQRODLKSPSOVAHLRAVYLHTGKPYPCPTGCTFRPHLQTL 466
DB 404 -HQGVHTGKRPAPQCGRAF---SHSNLTQQLHTGERPRCYDCGAFKAVL 458
QY 467 KSHVRIHTGKPYHCDPGLHFNKSQL-----RLHROKNGAATNTKYH 511
DB 459 LSHRIHTGKPYHCDPGLHFNKSQL-----RLHROKNGAATNTKYH 507

RESULT 15
US-08-820-170A-10
Sequence 10, Application US/08820170A
Patent No. 5831058
GENERAL INFORMATION:
APPLICANT: Isutomu, FUJIMARA
APPLICANT: Takechi, WATANABE
APPLICANT: Masato, HORIE
APPLICANT: Toyomasa, KATAGIRI
TITLE OF INVENTION: HUMAN GENE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sughrue, Mion, Zimm, Macpeak & Seas
STREET: 2100 Pennsylvania Avenue, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: United States
ZIP: 20037-3202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/820,170A
FILING DATE:
CLASSIFICATION: 536
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 293-7060
TELEFAX: (202) 293-7860
TELEX: 6491103
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 711 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-820-170A-10

Query Match 12.84; Score 357.5; DB 2; Length 711;
Best Local Similarity 32.04; Pred. No. 1,5e-19;
Matches 90; Conservative 29; Mismatches 113; Indels 49; Gaps 9;

QY 264 VQKCG-APASTPYLLTQAQDTSGSPSEPARPLPGSEFPSCONCEAVAGSSGLD---S 319
DB 448 VCIECGAFLQKAHLIVHQSHTEKEP-----YQCHNCGKSPISKSLDIHHR 495
QY 320 LVPDDEKPYKCOLCSSPFYKGNLASHRTVHTGEKPYHSGICGARFNRPNLKTSHRTH 379
DB 496 IHTG--EKPYECSDCGKTFTQKSHLNIHQKIHGERHVCSECKAFNOKSLISMHORIH 553
QY 380 SGEKPYKCECTGSRFVQVRSQPPSGFGCKEPARGVGQKGGFCSSORODLKSPS---QY 435
DB 554 TGEKPYKCECKGAFPS-----KSGFK-EHQRIHTGEKPYVCTECGKAFNGRSNFKHQI 607
QY 436 AHR-----AHVLIHTGEKPYPCPTGTRFRHLQTLKSHVRIHTG 475
DB 608 THTREPRFVCKGKAFVQKSELITHQRTMGEKPYECLDCGKSPSKPOLKVQRHITG 667
QY 476 EKPYHCDPGGLHFRHKSQQLRLHROKGAATNTKHYHILG 516
DB 668 ERPYVCECGKAFNRSNFKH-QTHTRDKSYKCSYSVKG 707

Search completed: July 14, 2003, 18:24:31
Job time: 8.81542 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:15:27 ; Search time 12.144 Seconds
(without alignments)
4965.877 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789

Sequence: 1 MGSPAPEGALGVREFTFRH.....RQKGNATNTKVYHILGSP 518

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 445758 seqs, 116419773 residues

Total number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2789	100.0	518	9 US-09-815-379-4	Sequence 4, Appl1
2	1095.5	39.3	706	10 US-09-107-058-2	Sequence 2, Appl1
3	1095.5	39.3	706	10 US-09-761-117-2	Sequence 2, Appl1
4	469.5	16.8	711	9 US-10-034-934-90	Sequence 90, Appl1
5	457.5	16.4	641	9 US-10-164-359-17	Sequence 17, Appl1
6	428	15.3	516	9 US-09-974-298-118	Sequence 118, App
7	426.5	15.3	516	9 US-09-764-891-4337	Sequence 4337, App
8	419.5	15.0	488	10 US-09-840-787-17	Sequence 17, Appl1
9	417	15.0	406	10 US-09-764-864-976	Sequence 976, Appl1
10	413	14.8	252	10 US-09-815-668A-12	Sequence 12, Appl1
11	412	14.8	468	10 US-09-864-761-37991	Sequence 37991, A
12	405.5	14.5	481	10 US-09-864-761-37651	Sequence 37651, A
13	399	14.3	241	10 US-09-764-864-935	Sequence 935, App
14	397	14.2	278	10 US-09-925-300-1323	Sequence 1323, App
15	394.5	14.1	492	10 US-09-864-761-42897	Sequence 42897, A
16	390.5	14.0	551	10 US-09-864-761-37065	Sequence 37065, A
17	390.5	14.0	610	9 US-09-989-920-212	Sequence 212, App
18	380	14.0	393	9 US-10-057-408-12	Sequence 12, Appl1
19	388	13.9	910	9 US-09-908-153B-40	Sequence 40, Appl1

20	388	13.9	922	9 US-09-908-153B-42	Sequence 42, Appl1
21	386.5	13.9	547	10 US-09-764-864-1065	Sequence 1065, App
22	384.5	13.8	439	10 US-09-864-761-33476	Sequence 33476, A
23	384	13.8	924	9 US-10-007-280A-231	Sequence 231, App
24	383	13.7	184	9 US-09-765-555-38	Sequence 38, Appl1
25	383	13.7	184	9 US-09-765-555-75	Sequence 75, Appl1
26	383	13.7	186	9 US-09-908-153B-36	Sequence 36, Appl1
27	383	13.7	525	10 US-09-864-761-33653	Sequence 33653, A
28	382.5	13.7	333	10 US-09-764-864-958	Sequence 958, App
29	382	13.7	184	9 US-09-765-555-40	Sequence 40, Appl1
30	382	13.7	359	10 US-09-764-864-996	Sequence 996, App
31	381.5	13.7	433	9 US-10-171-311-232	Sequence 232, App
32	381.5	13.7	717	9 US-09-881-578-2	Sequence 2, Appl1
33	380	13.6	180	9 US-10-106-698-5544	Sequence 5544, App
34	379	13.6	184	9 US-09-765-555-42	Sequence 42, Appl1
35	378	13.6	184	9 US-09-765-555-39	Sequence 39, Appl1
36	376.5	13.5	284	10 US-09-864-761-36268	Sequence 36268, A
37	376.5	13.5	573	9 US-10-157-031-36	Sequence 36, Appl1
38	376	13.5	284	10 US-09-764-864-932	Sequence 932, App
39	376	13.5	361	10 US-09-764-864-893	Sequence 893, App
40	375.5	13.5	338	10 US-09-864-761-46977	Sequence 46977, A
41	374.5	13.4	281	10 US-09-764-864-1076	Sequence 1076, App
42	374	13.4	196	10 US-09-764-864-991	Sequence 991, App
43	373.5	13.4	375	10 US-09-864-761-36336	Sequence 36336, A
44	373	13.4	184	9 US-09-765-555-41	Sequence 41, Appl1
45	372.5	13.4	614	10 US-09-864-761-43107	Sequence 43107, A

ALIGNMENTS

RESULT 1
US-09-815-379-4
Sequence 4, Application US/09815379
Publication No. US20030073613A1
GENERAL INFORMATION:
APPLICANT: RASTELLI, LUCA
APPLICANT: GERITSEN, MARY
TITLE OF INVENTION: ANGIOGENESIS ASSOCIATED PROTEINS AND NUCLEIC ACIDS
TITLE OF INVENTION: ENCODING THE SAME
FILE REFERENCE: 10716/35
CURRENT APPLICATION NUMBER: US/09/815.379
PRIOR APPLICATION NUMBER: 2001-03-22
PRIOR FILING DATE: 2000-03-22
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 4
LENGTH: 518
TYPE: PRT
ORGANISM: Homo sapiens
US-09-815-379-4

Query Match	100.0%	Score 2789	DB 9	Length 518
Best Local Similarity	100.0%	Pred. No. 6.3e-148		
Matches 518	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MGSPAPEGALGVREFTFRHSSDVIGNNELRRLGILTDVTLVGGQFLRAKAVLIACS	60	
DB	1	MGSPAPEGALGVREFTFRHSSDVIGNNELRRLGILTDVTLVGGQFLRAKAVLIACS	60	
QY	61	GFPTIFPGRAVGVDTLSLPGCPARGPAPLLDMYTSRLSLSPATPAVLAATYLIOM	120	
DB	61	GFPTIFPGRAVGVDTLSLPGCPARGPAPLLDMYTSRLSLSPATPAVLAATYLIOM	120	
QY	121	EHVQACHRFIOASYEPLGISLRPLEAPPTPTAPPGSPRRSGHPPTBSKSCGQ	180	
DB	121	EHVQACHRFIOASYEPLGISLRPLEAPPTPTAPPGSPRRSGHPPTBSKSCGQ	180	
QY	181	PPSPAPPPKACNMKKRYTYLVNSQASQAGSLVGRSSGQPCQARLPSGDEASSSSSS	240	
DB	181	PPSPAPPPKACNMKKRYTYLVNSQASQAGSLVGRSSGQPCQARLPSGDEASSSSSS	240	

QY 241 SSSSSSEBPIPGOSRLSPTAAVQFKGAPASTPYLLTSOADTSSPSEBARPLPGSE 300
DB 241 SSSSSSEBPIPGOSRLSPTAAVQFKGAPASTPYLLTSOADTSSPSEBARPLPGSE 300
QY 301 PFSCQNEAAGSSGGLDVLVPGDEDEPKYKCOLCRSSFYKGNLASHRTYHTEKPYHCS 360
DB 301 PFSCQNEAAGSSGGLDVLVPGDEDEPKYKCOLCRSSFYKGNLASHRTYHTEKPYHCS 360
QY 361 ICARFRRPANTLTHSHIHGSEKPYKCECTGSRFVQVRSOPSPSGFQKPARGGVGOKGF 420
DB 361 ICARFRRPANTLTHSHIHGSEKPYKCECTGSRFVQVRSOPSPSGFQKPARGGVGOKGF 420
QY 421 CSSQRODLKSPSQVAHLRAHVLHTEGKPYPCPTCGTRFRHLQTLKSHVRIHTEKPYH 480
DB 421 CSSQRODLKSPSQVAHLRAHVLHTEGKPYPCPTCGTRFRHLQTLKSHVRIHTEKPYH 480
QY 481 CDBCGHLFRKKSQRLHLRQKGAAATYTKVHYHILGSP 518
DB 481 CDBCGHLFRKKSQRLHLRQKGAAATYTKVHYHILGSP 518

RESULT 2

US-09-107-058-2
Sequence 2, Application US/09107058
Patent No. US20010010922A1
GENERAL INFORMATION:
APPLICANT: Dalia-Favera, Riccardo
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC
TITLE OF INVENTION: LOCUS Dcl-6
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,058
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0576/43771-A-PCT-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 706 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-107-058-2

Query Match 39.3%; Score 1095.5; DB 10; Length 706;
Best Local Similarity 35.7%; Pred. No. 1.4e-53; Indels 271; Gaps 18;
Matches 264; Conservative 60; Mismatches 144;

QY 1 MGSPAAEGALGVYRETRHSDVLGNLNLRLRGILITDVLVGGQPLRAHRAVLIACS 60
DB 1 MASPA--DSCI-----QTRHARVLLNLNLRSDILTDVIVVSSBQPAHKTIVLMAR 54
QY 61 GFPYSIFRGAGVGVVDVLSLPGEPEARGFAPLIDFMYTSRLSPATAPAVLAATYLOM 120
DB 61 GFPYSIFRGAGVGVVDVLSLPGEPEARGFAPLIDFMYTSRLSPATAPAVLAATYLOM 120

DB 55 GFPYSIFTDQLKCNLSVINLDEINPEGFCILLDFMYTSRLNLREGNIMAVNATAYLOM 114
QY 121 BHVYQACHRFIOAS----- 134
DB 115 BHVDTCKRFIOASEMVSALKPREEFLNSRMLMPQIMAYRGREVENNLPLRSABG 174
QY 135 -----YEPGLIS----- 141
DB 175 CSBAPAPSLYGLSTPPASYSYSHLPVSLFSDPEFQVDMVMPANPPKERALPCDS 234
QY 142 -----LRL----- 145
DB 235 ARPVPGEYRPTLEVS PNVCHSNISPKETIDPEARSDMHSVAGLKAAPASARNAPYF 294
QY 146 -----EAEPT-----PPTAP-----PGSPRSEGHDPPTES-----RSC 177
DB 295 PCDKASKERERSSEDELALHFEPPAPLRKGLVPSPOSDDCPNPTLACSSKAC 354
QY 178 ---SGPPSPASPPEKACNMKKYKIVLNS-----QASQAGSLVGRSSGOQCP 223
DB 355 ILQSGSPPAKSPDPAKCSMKYKFTVLNSLNQNAKPGCPQOAEIGRLSPRAYTAPAC 414
QY 224 QARLPBGDEASSSSSSSSSEBPIPGOSRL-----SPTATVQF----- 266
DB 415 QPPM-EPENLDQSPKLSASGEDSTIP-QASRLNNIVNRSMTGSRSSSESHSPLWHP 472
QY 267 ---KCG--APASTPYLTSQ---AODTSSPSEBARPLPSEFPSCQNEAAGSSG 316
DB 473 PKCTSCGSSQSPQAEKCLHTAGPTFAEMGEFQSEYSDSCENGAFPCNECCRSSEBS 532
QY 317 LD-SLVPGDEDEPKYKCOLCRSSFYKGNLASHRTYHTEKPYHCSICGARFRRPANTLTH 375
DB 533 LKRHTLQTHSDKPYKCDRCQASFRYKGNLASHRTYHTEKPYHCSICGARFRRPANTLTH 592
QY 376 STHSGEKYKCECTGSRFVQVRSOPSPSGFQKPARGGVGOKGFPCSSRODLKSPSPGY 435
DB 593 TTHSGEKYKCECTGARFV-----QV 614
QY 436 AHLRAHVLHTEGKPYPCPTCGTRFRHLQTLKSHVRIHTEKPYHCDPGLHFRKKSQRL 495
DB 615 AHLRAHVLHTEGKPYPCPTCGTRFRHLQTLKSHVRIHTEKPYHCEKCNLHFRKKSQRL 674
QY 496 LHLRQKGAAATYTKVHYH 514
DB 675 LHLRQKGAAITNTKVQYRV 693

RESULT 3

US-09-761-117-2
Sequence 2, Application US/09761117
Patent No. US20010012887A1
GENERAL INFORMATION:
APPLICANT: Dalia-Favera, Riccardo
TITLE OF INVENTION: CLONING AND USES OF THE GENETIC LOCUS
TITLE OF INVENTION: bcl-6
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/761,117
FILING DATE: 16-Jan-2001
CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:
 NAME: White, John P.
 REGISTRATION NUMBER: 28,678
 REFERENCE/DOCKET NUMBER: 0575/43771-A-PCT-US-X
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 278-0400
 TELEFAX: (212) 391-0525
 TELEX: 422523 COOP UI
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 706 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-09-761-117-2

Query Match 39.3%; Score 1095.5; DB 10; Length 706;

Best Local Similarity 35.7%; Pred. No. 1.4e-53;
 Matches 264; Conservative 60; Mismatches 144; Indels 271; Gaps 18;

1 MGSPAPBEGALGVREFTHSSDVLGNLNLRLGILTVTLVGGQPLRAKAVLIACS 60
 1 MASPA--DSCI-----QFTRHARDVLNLRLSRDLITDVIVVSRQFRAHKTIVLAWR 54
 61 GEFYSIFRGAGVGVVLSLPGGPEARGAPLLDFMYTSHRLSPATAVAVLAATYLOM 120
 55 GLPFYSIFTLQKCNLSVINLDPINDEGFCILDFYTSFLNREGNIMAVATANYLOM 114
 121 EHVYQACHRFIOAS----- 134
 115 EHVVDCKRFIRASBAEMVSAIKPREEFNLNMLPDIIMAYRGREVENNLPLRSADG 174
 135 -----YEPFGIS----- 141
 175 CSRAFAFLYGLSTPPASYSMTSHLPVSSLLFDEEFDVMPVAPNPFKERALPCDS 234
 142 -----LRPL----- 145
 235 ARPVEGSRPTLEVS PNVCHSNISPKETIPREARSDMHYSVABELKRAAPARAPAYF 294
 146 -----EAPPT-----PPTAP-----PPGSPRSEGHDPPTES---RSC 177
 295 PCDKASKEEERSSSEDEIALHFEPPNAPLNKGLVPSQSPQKSDCQPNPTEKSSKNAC 354
 178 -----SGCPSPASPDCKACNMKKYKTIYVNS-----QASQAGSLVGERSSGQPCP 223
 355 ILQSGSPSPAKSPPTDCKACSWKKYKFIYVNSLNQNAKPGGPEQALGRLSPRAYTAPAC 414
 224 QARLPFGDEASSSSSSSSSSSEEGIPGQSRP-----SPTAATVQF----- 266
 415 QPFW--EPENLDLOSPTKLSASGEDSTIP--QASRLNINIYVRSMTGSRSSSESPILYMP 472
 267 ----KCG--APASTPYLLTSQ-----AQTSGSPSEPARPLPGSEFPSCNCEAVACSSG 316
 473 PCTSGSGSPGHAEWCLHTAGPTFAEEMGETQSEYSDSSCEGAFPCNCDORFSEAS 532
 317 LD--SLVPGDEDPKYKQLCRSSFRYKGNLASHTVTYTGKRPYKCSICGARPNRPAULKH 375
 533 LKRNHTLQTHSKDKPKCDRCOASFRYKGNLASHTVTYTGKRPYKCNICGAFNPAULKH 592
 376 SRHSSEKPYKCEICGSRFVQVARSPPSGFGKPARGVQKKGFCSSQRODLKSPSPQV 435
 593 TRHSSEKPYKCEICGARFY----- 614
 436 AHLRAVLIHTGKRPYKCPCTGTRFRHLQTLKSHVRIHTGKRPYKCDPCGLHFRHKSQR 495
 615 AHLRAVLIHTGKRPYKCPCTGTRFRHLQTLKSHVRIHTGKRPYKCDPCGLHFRHKSQR 674
 496 LHLRQGAATNTKAYYHI 514
 675 LHLRQGAATNTKAYYRV 693

RESULT 4

US-10-034-934-90
 Sequence 90, Application US/10034934
 Publication No. US20030068624A1

GENERAL INFORMATION:
 APPLICANT: Recipon, Herve
 APPLICANT: Sun, Yongming
 APPLICANT: Chen, Sei-Yu
 APPLICANT: Liu, Chenghua

APPLICANT: Turner, Leah
 TITLE OF INVENTION: Compositions and Methods Relating to Lung Specific
 TITLE OF INVENTION: Genes and Proteins
 FILE REFERENCE: DEX-0245

CURRENT APPLICATION NUMBER: US/10/034,934
 PRIOR FILING DATE: 2000-10-26
 PRIOR FILING DATE: 2000-10-26
 PRIOR FILING DATE: 2000-10-26
 PRIOR FILING DATE: 2000-11-20
 PRIOR FILING DATE: 2000-11-20
 PRIOR FILING DATE: 2000-11-22
 NUMBER OF SEQ ID NOS: 142
 SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 90
 LENGTH: 711
 TYPE: PRT
 ORGANISM: Homo sapiens

US-10-034-934-90
 Query Match 16.8%; Score 469.5; DB 9; Length 711;
 Best Local Similarity 26.0%; Pred. No. 9.7e-19;
 Matches 162; Conservative 67; Mismatches 220; Indels 173; Gaps 22;

23 DVLGNLNLRLGILTVTLVGGQPLR---AHKAVLIACGFFYSIFRGAGV-GUDV 77
 16 NLHMEHMLRLGILCDVTVSVEYQVGRDFMAHKAVALAATSKFKEVFLNKSVDGRT 75
 78 LSLPGPEARGAPLLDFMYTSHRLSPATAVAVLAATYLOMENVYQACHRFIOASYEP 137
 76 NYVLNEVOVADPASTLEFVYTAQVVEEDRVQMLEVAEKLCLDLSCTFOLKQMLBS 135
 138 LGISLRPL-----EAPPT-----PPTAPPPGSPRSEGHDPPTES---RSCSGQPS 184
 136 VLLELQNTSESGEVEVSSGSQVSAAPARASVATDGPSPGTLTDLVYGERASNGMSD 195
 185 ASPDPACNMKKYKTIY----- 213
 196 LPPKSKDKLDCKKEVVPYKIRASGRLAGRLVLEIPKKYTRSLREQQTAEGCV 255
 214 GERSSGQPCQPARLDSG-----DEASSSSSSSSSSSEEGPIPCP 253
 256 GD-----YPCPDQSPDRVGTMEQVSKNECCQAGALELBSKKAPBESSEEDDEGE 311
 254 QSRLEPATAVQFC-----GAPASTPYLLTSQAQDTSGSP--- 289
 312 KKKSN-----FKSICEKAFLEYKSPFLKSKRHHGVADEVVNC-----DTCCGTAN 359
 290 ----SERAPPLGSR-FPSCNCG-----EAVAGCSGGLSLVPGDEDPKYKQL 333
 360 RCNLKSHORHVASSSRHPPCBLGCKKPKRKDKVKNVLAQVHEG-----GGF--RHRQ 411
 334 CRSSFRYKGNLASHTVTYTGKRPYKCSICGARPNRPAULKHTRHSSEKPYKCEICGSR 393
 412 CGKGLSSKTALRLHRTHTHTGDRPYCTEGCARFSPSLKTHMRHTGSKRPVCECGAR 471
 334 FVQVARSPPSGFGKPARGVQKKGFCSS---SORDLK-----SP----- 432
 472 FTQ-----NMHLIYHKKCHTGERPFMCETGSKPSASKEVLYKHNHRIHTGSKRPFCEVCF 525
 433 ---SOVAILRAVLIHTGKRPYKCPCTGTRFRHLQTLKSHVRIHTGKRPYKCDPCGLHFR 489
 526 RTFAQRNLSYOHIKVHTGERPYCCDCKGQFQTLNALORHRIHTGSKRPFMCNACGRFTT 585

QY 490 HKSQRLHLRQKHGAATNTKVH 511
Db 586 DKSTLRHH-----TSIH 597

RESULT 5

US-10-164-359-17
; Sequence 17, Application US/10164359
; Publication No. US2003001276A1
; GENERAL INFORMATION:
; APPLICANT: Chin, Khew-Voon
; TITLE OF INVENTION: Nucleic Acid and Protein Expressed Thereby and Their Involvement
; FILE REFERENCE: 601-1-108US
; CURRENT APPLICATION NUMBER: US/10/164,359
; PRIOR FILING DATE: 2002-08-06
; PRIOR APPLICATION NUMBER: PCT/US00/33438
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 60/169,418
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 17
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-164-359-17

Query Match 16.4%; Score 457.5; DB 9; Length 641;
Best Local Similarity 25.0%; Pred. No. 4e-18;
Matches 168; Conservative 56; Mismatches 195; Indels 253; Gaps 25;

QY 7 PEGALGYREPTSHSDVGLNLELRLR-GILTDVTLVGGOPLRHAKAVLTCGSPFYS 65
Db 11 PSCC--YTYVSHSTHTEMLNINQKNGRCFCDVLRVDBSPRAHRAVLAACSYFYS 68
QY 66 IFRGAGVG-----VDV---LSLPGRP-----EARGFAPLIDFMYTSRLRL 103
Db 69 VFSAQIGDGGADGADGADGATGATGATGATGATGATGATGATGATGATGATGATGAT 128
QY 104 SPRTAFAVLAATYLLQMEHVQACHRTQNSYEPGLISLRPLABEPTPTAP-----P 157
Db 129 RLSEFPELMTAAFLMLRSVTEICQEVTKQS-----NVQILVPPARADIMLFR 177
QY 158 PGSPRSEGHDPPTESRSC---SQGPPSPAPDPKAKMKKYKYLANSQASQASVLC 214
Db 178 PGT--SDLEPPLMTGALANANSNGIAGMPEBEA-----AAAGAAIA 221
QY 215 ERSSGQPCPGA-RLPSGDEASS--SSSSSSSSSEBGP1PGPOSRLSPTAATVQPCGA 270
Db 222 GQASLPVLPVDRPLPMVAGPLSPQLLTPSPSVASSAPPLTGKRGGRPRKAN----- 274
QY 271 PASTPVLTTSQADDTGSGPS--ERARPLP-----GSEFPGSCQC---EAVAGCS--GLD 318
Db 275 -----LDSM---FGSPGLREAGILPGCGKVFYTDANRLRQHAHQVTSIQIGYI 324
QY 319 SLVP-----GDEDEPKY-----COLCRSPRYKGNLASHRTVVTGKPYH 358
Db 325 DLPPPLGNGELIISDPDGPGRKSRTRKQVACEICGKIFRDVYHLNRKLSISGKRPIS 384
QY 359 CSICGARFNRPANLKTSHRIHSGE--KPYKCTGSRFVQVRSQPSGFGCKPARGVGQ 416
Db 385 CPVCGRFRKRDMSYVHRSHDSVGKPYICSGCGKF----- 422
QY 417 KGFCSGQRODLKSPSQVLAHAHV-LHTGKPYPCPFCGTRFPHLQTLKSHVRIH-- 473
Db 423 -----SRPDLNGHIKQVITSERPHKQPCYNASFATRDRLRSLACHED 466
QY 474 ----- 473
Db 467 KVPQCVGKTLRAAYADHLKHSBGSNFCISINREGQCKSHQDPIESSDSYIGDLSDS 526

QY 474 -----TGKPYHCDPCGLHFRHKSQRLHLRQKHGAAT 506
Db 527 DLKTEKQANGSPSCDMVAPRKMKMSDEKCYPCBEGSPFRSKSYLNKHQI----- 579
QY 507 NTKVHYHILGSP 518
Db 580 --KVHVRLGSP 589

RESULT 6

US-09-974-298-118
; Sequence 118, Application US/09974298
; Patent No. US20020156263A1
; GENERAL INFORMATION:
; APPLICANT: Chen, Huei-Mei
; TITLE OF INVENTION: GENES EXPRESSED IN BREAST CANCER
; FILE REFERENCE: PA-0037 P
; CURRENT APPLICATION NUMBER: US/09/974,298
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: 60/238,331
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 118
; LENGTH: 516
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020156263A1 5665139CD1
US-09-974-298-118

Query Match 15.3%; Score 428; DB 9; Length 516;
Best Local Similarity 34.6%; Pred. No. 1.4e-16;
Matches 110; Conservative 35; Mismatches 111; Indels 62; Gaps 12;

QY 223 PQARLP-----SGDEASSSSSSSSSSSEBGP1PGPOSRLSPTAAT--VOPK 267
Db 21 POSHLPELQFQAEKGYKDYDMERSVNSS-----LVSPQRISSVTXHSHTYE 72
QY 268 CGAPASTPVLTTSQADDTGSGP-----SERARPL-PSSEF-----PSCQNCBAV 310
Db 73 CNFVDS--LFTQKEKANIGTEHYKCNERKAFHOGILHTTIOIHTKETQKDCIGKI 129
QY 311 ACCSSGLDS--LVGDEDEPKYCOLCRSSFRYKGNLASHRTVVTGKPYHCSIGAREN 367
Db 130 FFKKSNLASHQRIHNG--EKPYKNECGKVFNHMSHLAQHRIHTGKPYKNECGKVFN 187
QY 368 PRANLKTSHRIHSGKPYKCTGSRFVQVRSQPSGFGCKPARGVGQKGFCSQROD 427
Db 188 QISHLAQHRIHTGKPYKNECGKVFNHMSHLAQHRIHTGKPYKNECGKVFN 241
QY 428 LKSPSQVLAHAHVLIHTGKPYPCPFCGTRFPHLQTLKSHVRIHTGKPYHCDPCGLH 487
Db 242 F-----SRNSYLVQHLIHTGKPYKNECGKVFNHMSHLAQHRIHTGKPYKNECGKV 297
QY 488 FRHKSQ-----RLHLRQK 501
Db 298 FSHKSLVNMHRIHTGK 315

RESULT 7

US-09-764-891-4337
; Sequence 4337, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIOR FILING DATE: 2001-01-17
; OTHER INFORMATION: consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231

SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 4337
LENGTH: 127
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (119)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (120)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
NAME/KEY: SITE
LOCATION: (127)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-891-4337

Query Match 15.3%; Score 426.5; DB 9; Length 127;
Best Local Similarity 81.7%; Pred. No. 4e-17;
Matches 89; Conservative 5; Mismatches 12; Indels 3; Gaps 2;

QY 1 MESPAPBAGALVYBFTTHSSDVIGNLNLRGLTDTVTLLVGGQPLRAKAVLIACS 60
DB 16 MESPAPBAGALVYBFTTHSSDVIGNLNLRGLTDTVTLLVGGQPLRAKAVLIACS 75
QY 61 GFFYSIFRGRAGVGVVLSLPGGPARGAPLIDFWYTSRLRLSPATAP 109
DB 76 GFFYSIFRGRAGVGVVLSLPGVPK-REASPL-VYTSCLRAACASXPP 121

RESULT 8
US-09-840-787-17
Sequence 17, Application US/09840787
Patent No. US20020058264A1
GENERAL INFORMATION:
APPLICANT: Lai, Preeti
Hillman, Jennifer L.
Bandman, Olga
Shah, Purvi
Au-Young, Janice
Yue, Henry
Guegler, Karl J.
Corley, Neil C.
TITLE OF INVENTION: HUMAN REGULATORY MOLECULES
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/840,787
FILING DATE: 23-Apr-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/518,865
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0356 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 17:

SEQUENCE CHARACTERISTICS:
LENGTH: 488 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRSTNOT03
CLONE: 641127
SEQUENCE DESCRIPTION: SEQ ID NO: 17 :
US-09-840-787-17

Query Match 15.0%; Score 419.5; DB 10; Length 488;
Best Local Similarity 29.5%; Pred. No. 4e-16;
Matches 110; Conservative 36; Mismatches 132; Indels 95; Gaps 12;

QY 176 SCSSGPPSPAPDPKACN-----KKKKYIYLNLSA-----SQASLVGRSS 218
DB 137 TCQNG-QLKESLDPIDCKCKDIHGMK--SQVSCSQRGHTBEKPCDHNNCKIINTSPD 193
QY 219 GQPCQARLPSCDEASSSSSSSSSSEGGPIPGQSRSLPTAATVQKCAPASTPYLL 278
DB 194 GHPYKIHIAEKGQYSGCGKNFSGSE-----LL 223
QY 279 TSQADTSGSPERARPLPGESEFSCQNC-EAVACSSGLDVLVGEDKPYKCOLCRSS 337
DB 224 LHQDHTTEKP-----YKCEOCCKGFTRSSSLIHQAVHTDEKPYKCDKCGKG 271
QY 338 PPKYNLASHRVHTEGKPYHCSIGARFNRPANLKTSHRGKPYKCTGSRFVG- 396
DB 272 FTRSSSLIHAAVHTEGKPYKCDKCGKSGSKLHQRVHTGKPYCECGKMSFSQR 331
QY 397 ----VRSQPPSGFGKPARGVGVGKGFCSQRODLKSPPSQVAHLRAHLITGEKPY 452
DB 332 SNLHIGQVHTG--BRPYKCGCKGF-----SSSNLHIRCHITGEKPYQ 376
QY 453 CPTGTRPRHLQTLKSHVRITGEKPYHCDPCGLHFRHKSQI---RLHLRQK----- 501
DB 377 CYECGKGFSSQSDLIHHRVHTGEKPYHCGKCGKFSQSKLHQRVHTGKPYECSKC 436
QY 502 -HGAATNTKRVHY 513
DB 437 GKGFSSQSNLH 449

RESULT 9
US-09-764-864-976
Sequence 976, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 976
LENGTH: 406
TYPE: PRT
ORGANISM: Homo sapiens
US-09-764-864-976

Query Match 15.0%; Score 417; DB 10; Length 406;
Best Local Similarity 36.6%; Pred. No. 4.5e-16;
Matches 94; Conservative 26; Mismatches 105; Indels 32; Gaps 7;

QY 267 KCGAPAS-TPYLLTQADTSGSPSE-----RAPPDPSBFPSCONCEA 309
DB 77 ECGKCSNMASHIHWQXHTGKPYECPGCGAPSEKSRRLRQGTHTGKPYKDCDCX 136
QY 310 VAGCSSGL--DSLVPGBDKPYKCOLCRSSFRYGNLASHRTVHTGKPYHCSIGAR 366

Db 137 AFSAGSLRIHQTHTG--EKPECECKGSPYKSTILVHQTHTGKEPECECKGSP 194
 Qy 367 NRPAUKTHSRIRSGEPEYKCTGSRFVVRSPGSGFQKPARGVGQKGFSSQRQ 426
 Db 195 SHHSGLRNHRRTHTGERPYKDECGKAF-----KLMSGLR-KHRTHTGKPYKQCKG 248
 Qy 427 DLKSPSQVAHLRAVLIHTGKPYPCPTGTRPHLQTLKSHVRHTGKPYKPCGL 486
 Db 249 AF-----GQKSLGRHHRHTGKPYKNCHEGAFSGKSNLRVHHRHTHTGKPYQCECGK 304
 Qy 487 HERNKSLRLHLRQKHG 503
 Db 305 TFRKSNLRGHRHTHTG 321

RESULT 10
 US-09-816-669A-12
 ; Sequence 12, Application US/09816669A
 ; Patent No. US20020137019A1
 ; GENERAL INFORMATION:
 ; APPLICANT: GARABEDIAN, Michael
 ; APPLICANT: TANEJA, Samir
 ; APPLICANT: HITTELMAN, Adam
 ; APPLICANT: MARKIS, Steven
 ; TITLE OF INVENTION: METHOD FOR SCREENING TRANSCRIPTIONAL COREGULATORY PROTEINS OF
 ; TITLE OF INVENTION: TRANSCRIPTION FACTORS, AND ANDROGEN RECEPTOR TRANSCRIPTIONAL CO
 ; FILE REFERENCE: GARABEDIAN=1.1A
 ; CURRENT APPLICATION NUMBER: US/09/816,669A
 ; CURRENT FILING DATE: 2001-03-26
 ; PRIOR APPLICATION NUMBER: 60/225,618
 ; PRIOR FILING DATE: 2000-08-15
 ; PRIOR APPLICATION NUMBER: 60/191,768
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 12
 ; LENGTH: 252
 ; TYPE: PRT
 ; ORGANISM: Human
 ; US-09-816-669A-12

Query Match 14.8%; Score 413; DB 10; Length 252;
 Best Local Similarity 37.2%; Pred. No. 4,6e-16;
 Matches 99; Conservative 25; Mismatches 96; Indels 46; Gaps 8;
 Qy 265 QPCKGPASTRVLLTSGAQTSGSPSRARPLPSEFSGQNCENAVAGSSGL---DSL 321
 Db 1 EFQCG---KVFAQTSL-----AHMVRHTGKPYKNDGGRAPSDRSLLTFHOAH 49
 Qy 322 PGDEDRPYKCOLCRSSFRYKGNLASHRTVATGKPYKCSICGARFNRPAUKTHSRIRSG 381
 Db 50 TG--EKPYKCHGEGKVFNRHSYLATHRHTGKPYKNCHEGAFMSHSLTTHKVIHNS 107
 Qy 382 EKPYKCTGSRVYVRSPGSGFQKPARGVGQKGFSS-----SORDL----- 428
 Db 108 EKPYKNCQCGKVFQ-----NSHLNHRHTHTGKPYKNCHEGAFVYSSSLTTHQAH 161
 Qy 429 --KSP-----SOVAHLRAVLIHTGKPYPCPTGTRPHLQTLKSHVRHTGK 477
 Db 162 TGKKPYKNCHEGKVFQNTOMHLNHRHTHTGKPYKCTGCKAKRVNSLTTVAHHTGK 221
 Qy 478 PYHCDPCGLHFRKSGQLRLHLRQKHG 503
 Db 222 RYKNCQCGKVFRRSSNLSHHRHTG 247

RESULT 11
 US-09-864-761-37991
 ; Sequence 37991, Application US/09864761
 ; Patent No. US20020048763A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.

APPLICANT: Rank, David R.
 APPLICANT: Hanzel, David K.
 APPLICANT: Chen, Wensheng
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
 FILE REFERENCE: Aemica-X-1
 CURRENT APPLICATION NUMBER: US/09/864,761
 CURRENT FILING DATE: 2001-05-23
 PRIOR APPLICATION NUMBER: US 60/180,312
 PRIOR FILING DATE: 2000-02-04
 PRIOR APPLICATION NUMBER: US 60/207,456
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: US 09/632,366
 PRIOR FILING DATE: 2000-08-03
 PRIOR APPLICATION NUMBER: GB 24263,6
 PRIOR FILING DATE: 2000-10-04
 PRIOR APPLICATION NUMBER: US 60/236,359
 PRIOR FILING DATE: 2000-09-27
 PRIOR APPLICATION NUMBER: PCT/US01/00666
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00667
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00664
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00669
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00665
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00668
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00663
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00662
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00661
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/00670
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: US 60/234,687
 PRIOR FILING DATE: 2000-09-21
 PRIOR APPLICATION NUMBER: US 09/608,408
 PRIOR FILING DATE: 2000-06-30
 PRIOR APPLICATION NUMBER: US 09/774,203
 PRIOR FILING DATE: 2001-01-29
 NUMBER OF SEQ ID NOS: 49117
 SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
 SEQ ID NO 37991
 LENGTH: 468
 TYPE: PRT
 ORGANISM: Homo sapiens
 FEATURE:
 OTHER INFORMATION: MAP TO AC005261.1
 OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
 OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 4.7
 OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.7
 OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
 OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
 OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.9
 OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
 OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
 OTHER INFORMATION: SWISSPROT HIT: P52742, EVALU 0.00e+00
 OTHER INFORMATION: EST HUMAN HIT: AF150239.1, EVALU 1.00e-105
 OTHER INFORMATION: EST_HUMAN HIT: AF150239.1, EVALU 1.00e-105
 US-09-864-761-37991

Query Match 14.8%; Score 412; DB 10; Length 468;
 Best Local Similarity 29.3%; Pred. No. 9,9e-16;
 Matches 114; Conservative 21; Mismatches 158; Indels 96; Gaps 9;
 Qy 184 PASPPKACNMKRYTYIVLNSQASQAGSVGRSSGQPCPQARLSPGDEASSSSSSSSS 243
 Db 5 PRSTPTTCLALSGISGIFWGLTQAGS--GDSQLGQPDQ---DGFSEWQGERLRPGL 58

QY 244 SSEEPIFGPQSRSLPTAATVQFKCGAPASTPYLLTSCQADTSGSPSERARPLPGSE-FF 302
 DB 59 DSGKEKLPKMSPKHDGCTADSVCGRIIQDRAVSLGDDVHDSDHSGGNPVIQEEENF 118
 QY 303 SCQNCBAVAG-----CGSGDLSTVPGDBDKPYKQCLCRSPRRYGNLASHRTVHT 352
 DB 119 KCNECEKFNKKRLLAHNERIHSGV-----KYECECEGKXFKFSKSTYLLQHHMVT 169
 QY 353 GKPYHCISICGARFNRPAULKTHSRHISGEKPYKCTGSRFVQ----- 396
 DB 170 GKPYKMECGKAFNKKSHLTOHQRHISGEKPYKCECGKAFTHNSTFYLHNRSHTEKP 229
 QY 397 -VRSQPPSGFGCKPA-----RGVGQK----- 417
 DB 230 FVCKEKGKAFRDRPGFIRHYIHSGENPYCECGKVFKHSYLMWHQDTHGEKPYECS 269
 QY 418 ---GGFCSSQRDLKSPSQVAHLRAHVLHITGEKPYPCPTGTRPRHLQTLKSHVRIHT 474
 DB 290 ECGKAFCEB-----ALIHVYIHTGEKPECECEGKAFNHSYLLKRRQRIHT 337
 QY 475 GKPYHCDPCGLHFRHKSQRLHLRQKHG 503
 DB 338 GKPYVCEGKAFTHCSTFILHKAHTG 366

RESULT 12
 US-09-864-761-37651
 / Sequence 37651, Application US/09864761
 / Patent No. US2002048763A1
 / GENERAL INFORMATION:
 / APPLICANT: Penn, Sharon G.
 / APPLICANT: Rank, David R.
 / APPLICANT: Hanzel, David K.
 / APPLICANT: Chen, Wensheng
 / TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
 / FILE REFERENCE: Aemolca-X-1
 / CURRENT APPLICATION NUMBER: US/09/864,761
 / CURRENT FILING DATE: 2001-05-23
 / PRIOR APPLICATION NUMBER: US 60/180,312
 / PRIOR FILING DATE: 2000-02-04
 / PRIOR APPLICATION NUMBER: US 60/207,456
 / PRIOR FILING DATE: 2000-05-26
 / PRIOR APPLICATION NUMBER: US 09/632,366
 / PRIOR FILING DATE: 2000-08-03
 / PRIOR APPLICATION NUMBER: GB 24263.6
 / PRIOR FILING DATE: 2000-10-04
 / PRIOR APPLICATION NUMBER: US 60/236,359
 / PRIOR FILING DATE: 2000-09-27
 / PRIOR APPLICATION NUMBER: PCT/US01/00666
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00667
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00664
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00669
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00665
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00668
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00663
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00662
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00661
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: PCT/US01/00670
 / PRIOR FILING DATE: 2001-01-30
 / PRIOR APPLICATION NUMBER: US 60/234,687
 / PRIOR FILING DATE: 2000-09-21
 / PRIOR APPLICATION NUMBER: US 09/608,408

/ PRIOR FILING DATE: 2000-06-30
 / PRIOR APPLICATION NUMBER: US 09/774,203
 / PRIOR FILING DATE: 2001-01-29
 / NUMBER OF SEQ ID NOS: 49117
 / SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 / SEQ ID NO 37651
 / LENGTH: 481
 / TYPE: PRT
 / ORGANISM: Homo sapiens
 / FEATURE:
 / OTHER INFORMATION: MAP TO AC000378.1
 / OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 4.4
 / OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.2
 / OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.2
 / OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.1
 / OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.7
 / OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.7
 / OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.3
 / OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.6
 / OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.9
 / OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.7
 / OTHER INFORMATION: SWISSPROT HIT: O14628, EVALU0 0.00e+00
 / OTHER INFORMATION: EST_HUMAN HIT: AUI33219.1, EVALU0 0.00e+00
 US-09-864-761-37651

Query Match 14.5%; Score 405.5; DB 10; Length 481;
 Best Local Similarity 34.6%; Pred. No. 2.3e-15;
 Matches 98; Conservative 38; Mismatches 108; Indels 39; Gaps 9;

QY 220 QPCPARLPDGDDEASSSSSSSSSSSSSSSECPPIPGQSRSLPTAATVQFKCGAPASTPYLLT 279
 DB 204 KCKCY-----ECCSVFISCSLSNQMIAG--EKL-----KC---ETWY--- 240
 QY 280 SQAQDTSGSPSERARPLDGESEFSCQNCBAVAGSSGDS---LVPGDBDKPYKCOLRS 336
 DB 241 -KGFNHSPVPSGQNEIGKPFKCEBCDSIFKPFSDITKRRITG--EKPYKCECGK 297
 QY 337 SFRRYGNLASHRTVITGEKPYHCISICGARFNRPAULKTHSRHISGEKPYKCTGSRFVQ 396
 DB 298 AVTQSHLSHRIRITGEKPYQCECGKVFTHCSLSNKKTHSHSEKXYTCECGNITKQ 357
 QY 397 VRSQPPSGFGKPARAGVQKGGFCSSQRDLKSPSQVAHLRAHVLHITGEKPYPCPTC 456
 DB 358 LSDLT-----KHKTHTEKPYKCECGK----TQSSNLIYKRIHTGEKPYKCEB 407
 QY 457 GTRFRHLQTLKSHVLIHTEKPYHCDPCGLHFRHKSQRLHLR 499
 DB 408 GRVFWFSDITYKHTHTEKPYKCECGKQFQSSNLIYHVR 450

RESULT 13
 US-09-764-864-935
 / Sequence 935, Application US/09764864
 / Patent No. US20020132753A1
 / GENERAL INFORMATION:
 / APPLICANT: Rosen et al.
 / TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
 / FILE REFERENCE: PT223
 / CURRENT APPLICATION NUMBER: US/09/764,864
 / PRIOR FILING DATE: 2001-01-17
 / Prior application data removed - consult PALM or file wrapper
 / NUMBER OF SEQ ID NOS: 1792
 / SOFTWARE: PatentIn Ver. 2.0
 / SEQ ID NO 935
 / LENGTH: 241
 / TYPE: PRT
 / ORGANISM: Homo sapiens
 US-09-764-864-935

Query Match 14.3%; Score 399; DB 10; Length 241;
 Best Local Similarity 46.6%; Pred. No. 2.6e-15;
 Matches 81; Conservative 15; Mismatches 68; Indels 10; Gaps 3;

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:07:56 ; Search time 24.0314 Seconds
(without alignment)

4441.379 Million cell updates/sec

Title: US-09-815-379-4

Perfect score: 2789
Sequence: 1 MGSPAPGALGTYREFTRH.....RQKQIAATNTKVNHLIGP 518

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: SP_TREMBL_21: *
2: sp_archaea: *
3: sp_bacteria: *
4: sp_fungi: *
5: sp_human: *
6: sp_invertebrate: *
7: sp_mammal: *
8: sp_mhc: *
9: sp_organelle: *
10: sp_phage: *
11: sp_plant: *
12: sp_rhodent: *
13: sp_virus: *
14: sp_vertebrate: *
15: sp_unclassified: *
16: sp_rv1rus: *
17: sp_bacteriopl: *
18: sp_archaeopl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2272	81.5	474	11	088282 mus musculu
2	509.5	18.3	539	4	015156 O15156 homo sapien
3	509.5	18.3	539	4	096P2 Q96P2 homo sapien
4	482	17.3	610	4	09NCP7 Q9NCP7 homo sapien
5	478	17.1	610	11	099MD8 Q99MD8 mus musculu
6	472.5	16.9	582	11	09CKJ8 Q9CKJ8 mus musculu
7	469.5	16.8	711	4	09H116 Q9H116 homo sapien
8	457.5	16.4	641	11	09JMG9 Q9JMG9 mus musculu
9	457.5	16.4	641	11	09JMG9 Q9JMG9 mus musculu
10	447	16.0	687	4	09HBE1 Q9HBE1 mus sapien
11	442.5	15.9	703	11	09R161 Q9R161 mus musculu
12	442	15.8	537	4	09HBE2 Q9HBE2 mus sapien
13	441	15.8	470	11	09PVP8 Q9PVP8 xenopus lae
14	439.5	15.8	569	11	09QZ48 Q9QZ48 rattus norv
15	439.5	15.8	545	4	09H117 Q9H117 mus sapien
16	437.5	15.7	565	11	088939 Q88939 mus musculu

17	436.5	15.7	725	4	096S24 Q96S24 mus sapien
18	432.5	15.5	555	11	062518 Q62518 mus musculu
19	430.5	15.4	555	11	0923D1 Q923D1 mus musculu
20	428.5	15.4	616	4	09Y529 Q9Y529 homo sapien
21	428.5	15.4	673	4	08TAL4 Q8TAL4 homo sapien
22	426.5	15.3	499	4	096CL5 Q96CL5 homo sapien
23	426.5	15.3	579	11	09NV10 Q9NV10 mus musculu
24	425	15.2	488	4	096R00 Q96R00 mus sapien
25	423	15.2	488	4	09NV05 Q9NV05 homo sapien
26	423	15.2	488	4	096NV1 Q96NV1 mus musculu
27	422.5	15.1	390	11	091W94 Q91W94 mus musculu
28	422	15.1	395	4	043593 Q43593 mus sapien
29	420.5	15.1	673	11	09QVT4 Q9QVT4 mus sapien
30	420	15.1	607	11	09D4A7 Q9D4A7 mus sapien
31	419.5	15.0	637	4	096IR2 Q96IR2 mus sapien
32	419	15.0	516	4	096SV5 Q96SV5 mus musculu
33	419	15.0	531	11	09D5K8 Q9D5K8 mus sapien
34	418.5	15.0	488	4	08TAS5 Q8TAS5 mus sapien
35	418.5	15.0	584	4	095365 Q95365 mus sapien
36	418	15.0	665	4	08TDL0 Q8TDL0 homo sapien
37	417.5	15.0	569	4	09H961 Q9H961 mus musculu
38	417.5	15.0	569	4	09H966 Q9H966 mus sapien
39	417.5	15.0	701	4	09UG02 Q9UG02 homo sapien
40	417.5	15.0	701	4	09UG02 Q9UG02 homo sapien
41	416.5	14.9	744	11	062788 Q62788 mus musculu
42	415.5	14.9	480	11	061117 Q61117 mus sapien
43	415.5	14.9	686	4	09NR94 Q9NR94 mus sapien
44	415.5	14.9	803	4	096TB6 Q96TB6 mus sapien
45	415.5	14.9	874	11	09R163 Q9R163 mus musculu

ALIGNMENTS

RESULT 1	088282	PRELIMINARY;	PRT;	474 AA.
AC	088282;			
DT	01-NOV-1998 (TREMBLrel. 08, Created)			
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	BAZF.			
GN	BAZF.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=98298267; PubMed=9632807;			
RX	Okabe S., Fukuda T., Ishibashi K., Kojima S., Okada S., Hatano M.,			
RA	Ebara M., Saiho H., Tokuhisa T.;			
RT	"BAZF, a novel bcl6 homolog, functions as a transcriptional			
RT	repressor".			
RL	Mol. Cell. Biol. 18:4235-4244(1998).			
CC	-1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).			
DR	EMBL; AB011665; BAA31223.1; -			
DR	HSSP; P15822; 1B80.			
DR	MCD; MGI:1278332; Bazf.			
DR	InterPro; IPR000210; BTB_P02.			
DR	InterPro; IPR000822; Znf_C2H2.			
DR	Pfam; PF00651; BTB; 1.			
DR	Pfam; PF00096; ZF-C2H2; 5.			
DR	PRINTS; PRO0048; ZINC_FINGER.			
DR	SMART; SM00225; BTB; 1.			
DR	SMART; SM00355; ZNF_C2H2; 5.			
DR	PROSITE; PS50097; BTB; 1.			
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_5.			
DR	PROSITE; PS0157; ZINC_FINGER_C2H2_2; 5.			
KW	DNA-binding; Nuclear protein; Zinc-finger.			
SO	SEQUENCE 474 AA; 51339 MW; 94886AB265B018D8 CRC64;			
Query Match	81.5%;	Score 2272;	DB 11;	Length 474;

Best Local Similarity 83.4%, Pred. No. 5,9e-161;
Matches 432; Conservative 11; Mismatches 31; Indels 44; Gaps 3;

```

QY 1 MGSPAAPGALGYVREFTRHSSDVGNINELRLRGITLDTVLVGGQPLRAKAVLIACS 60
DB 1 MGSTAAFGALGYVREFTRHSSDVGNINELRLRGITLDTVLVGGQPLRAKAVLIACS 60
QY 61 GFFYSIFRGRAGVGVDTVLSPGPGARGFAPLDPMYTSRLSPATPAVLAATYLYOM 120
DB 61 GFFYSIFRGRAGVGVDTVLSPGPGARGFAPLDPMYTSRLSPATPAVLAATYLYOM 120
QY 121 EHVVOACGRFIOASTYEPGLISLRPLEAPPTPPAPRPSRSEGHDPPTESRSCSG 180
DB 121 EHVVOACGRFIOASTYEPGLISLRPLEAPPTPPAPRPSRSEGHDPPTESRSCSG 180
QY 181 SPSPASDPKACNMKKYKYLIVNSQASQSLVGRSSGQPCPARLPSPGDEASSSSSS 240
DB 181 SPSPASDPKACNMKKYKYLIVNSQASQSLVGRSSGQPCPARLPSPGDEASSSSSS 240
QY 241 SSSSSSEGPITPGQSRLSPTAATVQFKGAPASTPYLLTSQAQDTSGSPERARPLPGE 300
DB 241 SSSSSSEGPITPGQSRLSPTAATVQFKGAPASTPYLLTSQAQDTSGSPERARPLPGE 300
QY 301 FFSQCEAVAVGSSGGLSLVPGDEDPYKCOLCRSSFRYKGNLASRTVHTGKPYHCS 360
DB 295 FFSQCEAVAVGSSGGLSLVPGDEDPYKCOLCRSAFRYKGNLASRTVHTGKPYHCS 360
QY 361 ICGARFNRPAHLKTHSRHSGEKPKYKCTGSRFVQVRSQPSGFGKPARGVGVGGGF 420
DB 355 ICGARFNRPAHLKTHSRHSGEKPKYKCTGSRFVQVRSQPSGFGKPARGVGVGGGF 420
QY 421 CSSGRDLPKSPQVAVLRAHVLHTGKPYKCTGCTTRRHLOTLKSHVRIHTGKPYH 480
DB 390 -----QVALLRAHVLHTGKPYKCTGCTTRRHLOTLKSHVRIHTGKPYH 436
QY 481 CDPGCLHFRHKSQRLHLRQKGAATNTKYHYHILGSP 518
DB 437 CDPGCLHFRHKSQRLHLRQKGAATNTKYHYHILGSP 474

```

RESULT 2

015156 PRELIMINARY; PRT; 539 AA.

AC 015156;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Kruppel-related zinc finger protein hckrox.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NC NCB1_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN.FIBROBLAST;
RA MEDLINE=98036076; Pubmed=9370309;
RX Widom R.L., Cullis I., Lee J.Y., Korn J.H.;
RT "Cloning and characterization of hckrox, a transcriptional regulator
of extracellular matrix gene expression";
RL Gene 198:407-420(1997).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
DR EMBL; AF007833; AAC51847.1; -;
DR HSSP; P08047; 1SP2.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00651; BTB_1.
DR Pfam; PF00096; zf_C2H2; 4.
DR PRINTS; PR00048; ZINCINGER.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00355; Znf_C2H2; 4.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.

KW DNA-binding; Nuclear protein; Zinc-finger
SQ SEQUENCE 539 AA; 58218 MW; AE7C5D83C04D583F CRC64;

Query Match 18.3%; Score 509.5; DB 4; Length 539;
Best Local Similarity 28.3%; Pred. No. 6.7e-30;
Matches 160; Conservative 53; Mismatches 190; Indels 163; Gaps 18;

```

QY 1 MGSPAAPGALGYVREFTRHSSDVGNINELRLRGITLDTVLVGGQPLRAKAVLIACS 60
DB 1 MGSP--EDDLIGL--PDPHSSBLSCNEQROLGLCDLTTRTGQLEVTRRAVLACS 56
QY 61 GFFYSIF-----RGRAGVGVDTVLSPGPGARGFAPLDPMYTSRLRL 103
DB 57 HYKKLFTGGGGAVVWAGSGTATGAGAGVCELDLV--GPRALG--ALLEAYATLTLT 113
QY 104 SPPTAPAVLAATYLYOMEHVVOACGRFIOASTYEPGLISLRPLEAPPTP----- 152
DB 114 SGANMPAVYQAARLLLEPCVIAACMEILLQGS-----GL-----EAPSPDEDCERARQY 162
QY 153 ----PAPRPGSPRRSEGHDPPTESRSCSGQPPSPASDPKACNMKKYKYLIVNSQASQ 208
DB 163 LEAFATATASGVPNGEDSPQVPLP-----PPEPPPRPVARRSRKRAFLDTKAR 215
QY 209 ASGLVGRSSGQPCQARLPSPGDEASSSSSSSSSEGPITPGQSRLSPTAATVQFKC 268
DB 216 ANHLVPE----VPTVAHPLTYEEBEVAGRVSSGGS-----GGDSISPTGTASPE 265
QY 269 GAPASTPYLLTSQAAD-----TSGSPERARPLPGEFPGSQNC---EAVAGCS 315
DB 266 GQSYEPYEGEEBEELVYPAYGLAOGGP-----PLSPBELGSDDEDAIDPDLAAYLS 320
QY 316 -GLDSLVPG--DEDK-----PYKCOLCRSSFRYKGNLASRTVHTGKPYHCS 363
DB 321 LHQDLVLAGLSDQDLVYKRSQMPQECFVCHKIITGCKLPRMRTHTGKPRACEVCG 380
QY 364 AEFNRPAHLKTHSRHSGEKPKYKCTGSRFVQVRSQPSGFGKPARGVGVGGGFCS 423
DB 381 VFTFNDKLIKTHMR----- 394
QY 424 QRODLKSPQVAVLRAHVLHTGKPYKCTGCTTRRHLOTLKSHVRIHTGKPYHCDP 483
DB 395 -----KHTGERPSPCHPCARFLHSYDLKQNHMLHTGDRPYECHL 434
QY 484 GCLHFRHKSQRLHLRQKGAATNTKY 509
DB 435 CHRAFAKEDHLQRLHKQNCLEVRTR 460

```

RESULT 3

096EP2 PRELIMINARY; PRT; 539 AA.

AC 096EP2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Similar to kruppel-related zinc finger protein hckrox.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NC NCB1_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=UTERUS;
RA Strauberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC012070; AAH12070.1; -;
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF00096; zf_C2H2; 4.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.

KW DNA-binding; Zinc-finger.
SQ SEQUENCE 539 AA; 58026 MW; 8683616504B05B7E CRC64;

Query Match 18.3%; Score 509.5; DB 4; Length 539;
Best Local Similarity 28.3%; Pred. No. 6.7e-30; Indels 163; Gaps 18;
Matches 160; Conservative 53; Mismatch 190;

1 MGSPAPAGALGVREFTTHSSDVLGNLNLRLGLITVTLVGGQPLRAKAVLIACS 60
1 MGSP--EDDLIGI--PFPDHSSELISCLNEQRLGHLCDLITRTQGLERTRRAVLAACS 56
61 GFFYSIF-----RGRAGVGVLSLPGSPKARPAFLIDPMYTSRLRL 103
57 HYFKKLFTEGGGAVWAGSGSGTATGAGAGVCELDLV--GPEALG--ALLEFAYATLT 113
104 SPATAPVLAATYLTOMEHVQACHFIDASVEPGISLRPLEAEPT----- 152
114 SSANMPAVIAQARLEPCVIAQMEILQS--GL-----EAPSPDEDCERARQY 162
153 ----PTAPPSPRRSEGHDPPTESRSCSQPPSPASFPKACNMKKYKVIYVNSQSQ 208
163 LEAPATATSGVFNGBDSPQVLP-----PPPPRPVARRRKRKRAFLQTKGAR 215
209 AGSLVGRSSGQPCPOARLPDGDEASSSSSSSSSESGPIPGQSRLSPTAATVQFKC 268
216 ANHLVPE---VPTVAHPLVTEEBEVAAGRVSSGGS-----GPGDSYSPPTGASPE 265
269 GAPASTPVLLTQAOQ-----TSGSPSRAPPLGSEPFSCQNC---EAVAGCS 315
266 GQOSTEYEEGEEBELVYPRAYGLAQGGP---FLSPBELGSEDAIDPLMVLSS 320
316 -GLDGLVPG--DEDK-----PYKCOLCRSSFPYKKNLASHRTVHGEKPYHCSIG 363
321 LHQDNAPRLDQDQKLVKRKRSQMPQECPCVCHKIHGAKLPRHMTHTGKEKPFACEVGG 380
364 ARPNRPANIKTSRIHSGEKPKYCETCSRFVQVRSQPPSGQKPARGQVGKGFCS 423
381 VFETRNDKLKIHMR-----KHTGERPYSCPHCPARFLHSYDLKNHMLHTGDRPYECL 434
424 QNODLKSPPSQVAHLRAHVLHNGEKPYCPTCGTRFRHLQTLKSHVRIHTEKPYHCDP 483
395 -----KHTGERPYSCPHCPARFLHSYDLKNHMLHTGDRPYECL 434
484 CGLHPRHKSQRLRLHROKHAATNTK 509
435 CHKAPAKEDHLQRLHKGQNCLEVTR 460

RESULT 4
GNPC7 PRELIMINARY; PRT; 610 AA.
AC GNPC7; 01-OCT-2000 (TEMBLrel. 15, Created)
DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Zinc finger protein SBB121 (Myoneurin).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCB1_TaxID=9606;
RN NCB1_TaxID=9606;
RP SEQUENCE FROM N.A.
RA Zhang W., Wan T., Yuan Z., Cao X.;
RT "A ZINC finger protein SBB121 expressed in human dendritic cells";
RT Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
RN NCB1_TaxID=9606;
RP SEQUENCE FROM N.A.
RA MEDLINE=20334280; PubMed=10873615;
RA Alliel P.M., Seddigi N., Goudou D., Cifuentes-Diaz C., Romero N.,
RA Velasco E., Rieger F., Perin J.P.,
RT "Myoneurin, a novel member of the BTB/POZ-zinc finger family highly
RT expressed in human muscle";
RT Biochem. Biophys. Res. Commun. 273:385-391(2000).

[3]
RP SEQUENCE FROM N.A.
RC TISSUE=EMERYO;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H., Sugawara M.,
RA Wagaitsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA Niomiya K., Iwayanagi T.;
RT "NEO human cDNA sequencing project";
RL Submitted (AUG-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF076249; AAF80160.1; -
DR EMBL; AF148848; AAF73138.1; -
DR EMBL; AK021646; BAB13862.1; -
DR HSSP; P08048; 7ZNP.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00651; BTB; 1.
DR Pfam; PF00096; zf-C2H2; 8.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS50028; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
KW DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 610 AA; 68681 MW; 842B62DDC081480 CRC64;

Query Match 17.3%; Score 482; DB 4; Length 610;
Best Local Similarity 26.8%; Pred. No. 8.5e-28;
Matches 149; Conservative 65; Mismatch 212; Indels 130; Gaps 16;

16 EFTRSSDVLGNLNLRLGLITVTLVGGQPLRAKAVLIACGFFYSIRGRA--CY 73
2 QSHHCHEHLERLNQREAGFLCDCTIVIGEFQFAHNVLASFSRYCATYRSTSENNV 61
74 GVDVSLPGSPARFAPLDPMYTSRLSPATAPVLAATYLTOMEHVQACH----- 128
62 FLDDQGV---KADFQGLLEIYTGTLNLDSMNKELHQAADYKAVEVYTKKIKMED 117
129 -----RFTQASYEPLGISLRP-----LEAPPTPTAPP 159
118 FAFIANPSTETISITGNILNQTCCLTLRDYNNRKESEVSTDLIQNPQKALAKSS 177
160 -SPRSEGHDPPTESRSCSQP-----PSPASPPKACNMK---KY 197
178 QTKKKKKAFFNSPKTQONTVOYPSDILENASVELFLDANKLPTPVQVQAOINDSELEL 237
198 KYIVNSQASQ--AGSLVGRSSGQPCPOARLP--GDEASSSSSSSSSESGPIPG 252
238 TSVENTPPADIVHTVYVKKRGSQPNCLAKHSMSNINAVKSPYEAENGEE--LDQ 295
253 PQSRLSPTAATVQFKCAPASTPVLLTQAOPTSGSPSEARAP---LGSSEFSCQNC 308
256 RYSKAKPMONT---CGVVFSEASSLRHMRHKG---VPPYCHLCGKAFTQCNOLK 346
309 AVAGSSGGLSLVPEDEDKPYKCOLCRSSFRYKKNLASHRTVHTG--EKPYHCSIGARP 367
347 THVRHTG-----EKPYKCELCDGFAKCOLVHRSRHHHEEKPYKDVCLNLP 397
368 RPANIKTSRIHSGEKPKYCETCSRFVQVRSQPPSGQKPARGQVGQKGFCSQROD 427
398 TSSNLKIHARKKSGSKPYVCDRCGRF----- 424
428 LKSPSQVAHLRAHVLHNGEKPYCPTCGTRFRHLQTLKSHVRIHTEKPYHCDP 487
425 -----AQASTLYVHRHTGKEKPYVCDRCGKAFAVSSSLITHSRHGTGKEKPYICGICGS 479
488 FHKSQLHLRQKKG 503
480 FISSGELNKHFRSHTG 495

```

RESULT 5
099MD8 PRELIMINARY; PRT; 610 AA.
ID 099MD8;
AC 099MD8;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 17, Last annotation update)
DE Myonaurin.
GN MYN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RA Alliel P.M., Seddigi N., Goudou D., Clifuentes-Diaz C., Romero N.,
RA Velasco E., Rieger F., Perin J.P.;
RA Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
RT EMBL: AF349561; AK18605.1; -.
RL HSSP; P08048; 7ZNF.
DR MGD; MG1:1931415; Myrn.
DR InterPro; IPR000210; ZNF_C2H2.
DR InterPro; IPR000822; ZNF_C2H2.
DR Pfam; PF00051; BTB; 1.
DR Pfam; PF00096; ZF_C2H2; 8.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00355; Znf_C2H2; 8.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_1; 8.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 8.
DR DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 610 AA; 68520 MW; 0C356BE1E00731F CRC64;

Query Match 17.1%; Score 478; DB 11; Length 610;
Best Local Similarity 26.8%; Pred. No. 1.7e-27;
Matches 150; Conservative 59; Mismatches 213; Indels 138; Gaps 14;

16 EFTRHSSDVLGNLELRILRGILTVTLVGGPRLRAHKAVALACSGFFYSIFRGA--GV 73
2 QVSHCEHLERLNKQREAGFLDCIVVIGEPFKAHNRVLAFFSYFGAIVYSTENN 61
74 GUDVLSLPGGPEKRGAPLIDFNYTSLRLSPATAVAALATYLOMEHYVQAC----- 127
62 FLDSQGV---KADGFOKLEPIYTGTLNDWNVKEIHQADYLVVEEVTWKCKIKMED 117
128 HRFIQASYEPLGLSLPRLAEPPPTAPPPG-----PRSEGHDPPTSRSC 177
118 FAFI-----ASPESTISITIGITELNQACCLTLTDYNNRRESEVSTDS 162
178 SQGPPSPAPDPACAKWKYKXIVLNSQASQASLVERSSGQPCQARLPSCDEASSS 237
163 VQANPKPRALTKKSSQSKKKAFSSQKQCG-----SKAVQYSDVLSEASVLFLE 214
238 SSSSSSSSEEGPFGPGQS-----RLSPFAATV-----QFKCG----- 269
215 TSKLSSPPVVEQIIQGNDSSELTLTSVVENTFPQDIIVQTVYVKRKRSQSHCALKEHM 274
270 ---APASTPYLINSQADTSGSPSERARPL-----PSSEFSSQ 305
275 SNIASVKSFELEENAGEILDAPS-KAKPWCNTCGVFEASSLRRHMRHKGVKPYVCH 333

```

```

306 NC-EAVAGCSSGLDVLVPEDEDPYKCOLCRSSFRYKGNLASHRTVHTG-EXPHYCSICG 363
334 LCGKAFQCNQKHTVHTHTGERPYKCELCXGFAQKQCLVHSHMHGEEKPYKCDVCN 393
364 AEPNRPANLKTSHRHSGEKPYKCELCXGFAQKQCLVHSHMHGEEKPYKCDVCN 423
394 LQFAVSSNKLKHARHSGSKPYKCDRCGRF----- 424
424 QKQDLKSPSQAHLRAVYLHTGEKPYKCPCTGFRHLQTLKSHVRLHTGEKPYKCDP 483
425 -----AQSSTLYHVRHTGKPYKCDTCGAFVAVSSLLTHSRKHTGEKPYKCI 475
484 CGLHFRHKSQRLHURKKG 503
476 CGKSFISGELNKHFRSHTG 495

RESULT 6
099CXJ8 PRELIMINARY; PRT; 582 AA.
ID 099CXJ8;
AC 099CXJ8;
DT 01-JUN-2001 (TEMBLrel. 17, Created)
DT 01-JUN-2001 (TEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE 281001IC24Rik protein.
GN MYNN OR 281001IC24Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCB1_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC HEAD;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Ichi Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaoka I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsumoto Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stambli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzaletti J., Momberts P.,
RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohlschki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK014315; BABN29266.1; -.
DR HSSP; P08048; 7ZNF.
DR MGD; MG1:1931415; Myrn.
DR InterPro; IPR000210; ZNF_C2H2.
DR InterPro; IPR000822; ZNF_C2H2.
DR Pfam; PF00051; BTB; 1.
DR Pfam; PF00096; ZF_C2H2; 7.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00225; BTB; 1.
DR SMART; SM00355; Znf_C2H2; 7.
DR PROSITE; PS50097; BTB; 1.
DR PROSITE; PS50097; ZINC_FINGER_C2H2_1; 6.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
DR DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 582 AA; 65482 MW; 7415F8370CDDC6F8 CRC64;

Query Match 16.9%; Score 472.5; DB 11; Length 582;
Best Local Similarity 26.4%; Pred. No. 4.1e-27;
Matches 150; Conservative 63; Mismatches 208; Indels 147; Gaps 15;

```

```

QY 16 EFTRHSSDVLGNLNLRLGILTDVTLVVGQPLRAHKAVALIACSGFFYSIFGRGA--GV 73
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 2 QVSHHCHEHLERLNKREAGPLCDGVVIGEFQKARHNVTLASFYSFGAIIYRSTENN 61
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 74 GVDVLSLPGCFEAKGAPLLDFMYTSLRLSPATAVAVLAATYLOMENVQA----- 127
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 62 PLDQGV-----KADGFQKLEFLYGTGLNDSMNVKEIHCAADYLVVEEVTCKIKMED 117
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 128 HRFIOGVYPLGLSLARPLEAEPTPTAPPPG-----PRRSEGHDPPTSRSC 177
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 118 PAFI-----ASPSSTISITGIELNQO?CLITLDPYNNREKSEVSTGS 162
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 178 SCGPSPAPDPKACMKKYKYLVLNSQAQAGSLVGRSSGQPCQARLPSCDEASSS 237
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 163 VQANPQPALTLTKSSQSKKKKKGFSQKGGQ-----SKAVQYSDVLESASVLPLE 214
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 238 SSSSSSSSEEGPDPGQS-----RLSPATAV-----QFKCG----- 269
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 215 TSKLSPPVVEQIIOGNDSESELSTSVVENTFPQDIQVQTVYRKRRKKSQSHCALKEHSM 274
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 270 ---APASTPYLLTSQADOTSGSPSERAPL-----PSEFFSQ 305
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 275 SNIAVSKSPYLEMAGEELDAPPS-PAKPCNTCGVFSFASLRHMRHKGKVPYVGH 333
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 306 NC-EAVAGCGSLDGLVPGDEDEKPYKCOLCRSSFRYKGNLASHRTVHTG-EKPYHCSIG 363
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 334 LCGKATQCNQKLTHTVHTGERPYKCELDCKGFAKQCVLFRSHHHEKPYKCDVGN 393
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 364 AAFNRPAHLKTSRIHSGEKPYKCTCGSRFVQVRSQPRGFGKPARGVGQKGFCS 423
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 394 LQFATSNLKIHKRKSSEKPYVCDTCGKAFVAVSSLITSRKHTGERPICEL 424
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 424 QKQDLKSPSSQVAHLRAHVLHTEKPYPCPTCGTFRHLQTLKSHVRIHTEKPYHCDP 483
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 425 -----AQASTLYTHVRHRTGKPYVCDTCGKAFVAVSSLITSRKHTGERPICEL 475
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 484 CGLHFRHKSQRLRLHROKHGAATNTKVN 511
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 476 CGNSTYDINKLKH-----KTKVN 494
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

Query Match 16.8%; Score 469.5; DB 4; Length 711;
Best Local Similarity 26.0%; Pred. No. 8,6e-27;
Matches 162; Conservative 67; Mismatches 220; Indels 173; Gaps 22;

QY 23 DVLGNLNLRLGILTDVTLVVGQPLR---AHKAVLIACSGFFYSTIRGAGV-GVDV 77
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 16 NLHHEHLEHLRLGHLCDVTSVVEYQGVKDFMAHKAVALAATSKPFKEVPLNKSVDGTRT 75
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 78 LSLPGPEAKGAPLLDFMYTSLRLSPATAVAVLAATYLOMENVQA?CHRFIOGVY 137
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 76 NYTLNEVQVADPASPLEFVYTAQVVEEDRVORMLEVAKLCLDSEFTCFOLKQMLSS 135
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 138 LGISIRPL-----EAEPT---PPTAPPPSPRSEEG-HPDPTES-----RSCSQPSP 184
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 136 VLEELQNPESQEVLEVSSGSQVSAAPAPASVATGPHPSGLTDLDPYGERASNGMSSD 195
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 185 ASPDPKACMKKYKYLVLNSQAQAGSLVGRSSGQPCQARLPSCDEASSS 213
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 196 LPPKSKDLDKCKKEVVPYPKIRRASGLRAGKVFVFI PKKYYTFLRBOQKTAEGDV 255
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 214 GERSSGQPCQARLPSC-----DEASSSSSSSSSEEGPIPP 253
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 256 GD-----YRCPDQSDPDRVGTENEQVS?KQEGQAGALELSTKAGPEEBEEDDEEG 311
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 254 QSLSPATAVQFKC-----GAPASTPYLLTSQADOTSGSP--- 289
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 312 KKKSN-----FKSICEKAPLYEKSPLKSKHGHVATEVYVC-----DTCCGTAN 359
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 290 ---SERAPLPGBS-FPSCQNC-----BAVAGCGSLDGLVPGDEDEKPYKCOL 333
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 360 RCNLKSHQHIVASSERHFPCELCGKKPKKQVKGHVLQVHGG-----GGE--RHRQ 411
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 334 CASSRYKGNLASHRTVHTGERPYKCTCGARFAPRPAHLKTSRIHSGEKPYKCTCGSR 393
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 412 CCKGLSSKTAALHRTHTGDRPYCTCEGARFQSPSLKTMRIHTEKPYVCECGAR 471
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 394 FVQVRSQPRGFGKPARGVGQKGFCS-----SORODLK-----SP----- 432
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 472 FTQ-----NHLILYHKRCHTERPFCETCKGSAKSEYLYKHNRHITGSRPFKCEVCF 525
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 433 ---SQVAHLRAHVLHTEKPYPCPTCGTFRHLQTLKSHVRIHTEKPYHCDPGLHNR 489
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 526 RTFAQRNSLYGHIKHTGERPYCCDQCGKQFLQNALQHRHRIHTEKPYVCECGAR 585
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
QY 490 HKSQRLHLRQKHGAATNTKVN 511
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 586 DKSTLRRH-----TSIH 597
   :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

RESULT 7
09H16 PRELIMINARY; PRT; 711 AA.
ID 09H16;
AC 09H16;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE DJ322G13.2.3 (zinc finger protein FLJ21794, isoform 3).
GN DJ322G13.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCB1_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA Lovell J.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AL096677; CAC03438.2; -.
DR HSSP; P08047; ISP2.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF00651; BTB_1.
DR Pfam; PF00096; Zf-C2H2; 10.
DR ProDom; PD000003; Znf_C2H2; 1.
DR SMART; SM00225; BTB_1.
DR SMART; SM00355; Znf_C2H2; 10.
DR PROSITE; PS50097; BTB_1.
DR PROSITE; PS50028; ZINC_FINGER_C2H2_1; 9.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 10.
DR DNA-binding; Metal-binding; Zinc-finger.
KW SEQUENCE 711 AA; 80491 MW; 9209B850193BCF1A CRC64;

```

```

RESULT 8
09UDU0 PRELIMINARY; PRT; 641 AA.
ID 09UDU0;
AC 09UDU0;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE WUGSC:H.DV40WZ3.1 protein (Zinc finger sarcoma long A isoform)
DE (BTB-POZ domain zinc finger transcription factor).
GN ZSG OR R1A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
OX NCB1_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1107-1108(1998).
RN (2)
RP SEQUENCE FROM N.A.
RA Gattung S., Geisel C., Murray J.;

```

RT "The sequence of Homo sapiens PAC clone RP3-400N23.";
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RA Waterston R.H.;
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RA Waterston R.;
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20404868; PubMed=10949935;
 RA Mestrangeo T., Modena P., Tornelli S., Bullrich F., Testi A.,
 RA Mezzalana A., Radice P., Azarelli A., Piloti S., Croce C.,
 RA Pierotti M., Sozzi G.;
 RT "A novel zinc finger gene is fused to EMS in small round cell tumor.";
 RL Oncogene 19:3799-3804(2000).
 RN
 RP SEQUENCE FROM N.A.
 RA Chin K.-V., Yang W.-L., Kudoh K.;
 RT "Novel CAMP Signalling via the Regulatory Subunit of the CAMP-
 RT dependent Protein Kinase.";
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC005003; AF01349.1; -
 DR EMBL; AF254083; AAG09032.1; -
 DR EMBL; AY028384; AAK19024.1; -
 DR HSSP; Q05515; ICS3.
 DR TRANSFAC; T04797; -
 DR InterPro; IPR000637; AT_hook.
 DR InterPro; IPR000515; BPD_transp.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR000822; ZnF_C2H2.
 DR Pfam; PF02178; AT_hook.1.
 DR Pfam; PF00651; BTB.1.
 DR Pfam; PF00096; Zf-C2H2; 7.
 DR ProDom; PD000003; ZnF_C2H2; 1.
 DR SMART; SM00384; AT_hook.1.
 DR SMART; SM00225; BTB.1.
 DR PROSITE; PS00355; ZnF_C2H2; 7.
 DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
 DR PROSITE; PS00097; BTB.1.
 DR PROSITE; PS00354; HMGI_Y; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
 DR DNA-binding; Metal-binding; Zinc-finger;
 KW SEQUENCE 641 AA; 69079 MW; 2951535380FOCE CRC64;
 SQ
 Query Match 16.4%; Score 457.5; DB 4; Length 641;
 Best Local Similarity 25.0%; Pred. No. 6e-26;
 Matches 168; Conservative 56; Mismatches 195; Indels 253; Gaps 25;
 QY 7 PEGALGVREFTTHSSDVLGNLELRIR-GILTDVTLVGQPLRAHKAVALINGCFEFS 65
 DB 11 PSCC--YTVQVSHRSTELNINLQOKNGRFCDVLRVGDSEFPARHRAVLAACSEYFS 68
 QY 66 IFGRAGVG-----VDV---LSLPGP-----EARGFAPLIDFMYTSLRL 103
 DB 69 VFSAQIGDGGAAAGCPADVGATAPAGGAGGSSRELEMTISSKVGDIIDFYRTIRIV 128
 QY 104 SPATAVAVLAATYVLOWEHVQACHRFIOASYPELGISLRPLAEPTPTAP-----P 157
 DB 129 RLBSFELMTAAKFLMLRSVIEICQEVYKOS-----NVQILVPPARADIMLFRP 177
 QY 158 PGSPRSEGHPPDPPTESRSC---SQGPPSPASBPYACMKKKYKTVLNASQASGLVYG 214
 DB 178 PGT--SDLGFPDLMITGALAAANSNIGASMGQEEER-----AFAAGAAIA 221
 QY 215 ERSSGQPCPOA-RLPSGDEASS---SSSSSSSSSEGGPIPGQSRSLPPTATVQKCGA 270
 DB 222 GQASLVLVEQVDLPMVAGLSPQLITSPSPVASSAPLITGKRGGRPKAN----- 274
 QY 271 PASTPYLTISQADTSGSPS--ERARPLP---GSEFFSCQNC---EAVAGCSS---GLD 318

DB 275 -----LDSM-----FGSGGLREAGILPCGLCGKFTDANRLRQHEAQHGYTSLQLYI 324
 QY 319 SLVP-----GDEDKPYK-----COLCRSSFRRKGNLASRTVHTGEKPYH 358
 DB 325 DLPPEPLGNGPLISEDDPGPRKRSRTRKQVACIGKIFRVYHLNRLKLSHSGKEPS 384
 QY 359 CSICGARFNRPANLKTSHRSIHSGE--KPKCECTGSRFVQVNSOPSPGFGKPARAGVQ 416
 DB 385 CVVCGLRFRKXKRMGSHVTHSDGSGKPYICQSCGKF----- 422
 QY 417 KSGFCSSQRQDLSPPSOVAHLRAHV-LIHTEKPYPCPTCTTRPHLTLKSHVRIH-- 473
 DB 423 -----SRPDHLNGHIKQVHTSERPHKCOTCNASFATRDRLSLHACHED 466
 QY 474 ----- 473
 DB 467 KVPQVCGKYLRAYADHLKXHSQPSNFCISINREGQCSHODPIESSDSYGDLSDAS 526
 QY 474 -----TSEKPYHCDPCGLAFPRHKSQTLRLHROKHAAT 506
 DB 527 DKTPEKQANGSFQCDMAVPKMKHESDGEKKYPCPEGCSFRRSKSYLNKHIO----- 579
 QY 507 NTKVHYHLGSP 518
 DB 580 --KVHVRALGSP 589
 RESULT 9
 Q9JMG9 PRELIMINARY; PRT; 641 AA.
 ID Q9JMG9
 AC Q9JMG9;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Transcription factor MAZR.
 GN ZFP278 OR MAZR.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_Taxid=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=EMBRYO;
 RX MEDLINE=20136053; PubMed=10669750;
 RA Kobayashi A., Yamaguchi H., Hoshino H., Muto A., Sato K., Morita M.,
 RA Hayaishi N., Yamamoto M., Igarashi K.;
 RT "A combinatorial code for gene expression generated by transcription
 RT factor bach2 and MAZR (MAZ-related factor) through the BTB/POZ
 RT domain.";
 RL Mol. Cell. Biol. 20:1733-1746(2000).
 DR EMBL; AB029397; BAA90874.1; -
 DR HSSP; Q05516; ICS3.
 DR TRANSFAC; T04796; -
 DR MGD; MGI:1891832; Zfp278.
 DR InterPro; IPR000637; AT_hook.
 DR InterPro; IPR000515; BPD_transp.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR000822; ZnF_C2H2.
 DR Pfam; PF02178; AT_hook.1.
 DR Pfam; PF00651; BTB.1.
 DR ProDom; PD000003; ZnF_C2H2; 1.
 DR SMART; SM00384; AT_hook.1.
 DR SMART; SM00225; BTB.1.
 DR SMART; SM00355; ZnF_C2H2; 7.
 DR PROSITE; PS00402; BPD_TRANSF_INN_MEMBER; UNKNOWN_1.
 DR PROSITE; PS00097; BTB.1.
 DR PROSITE; PS00354; HMGI_Y; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
 DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 6.
 DR DNA-binding; Metal-binding; Zinc-finger;
 KW SEQUENCE 641 AA; 69138 MW; C4BEA972CC15A877 CRC64;
 SQ

Query Match 16.4%; Score 457.5; DB 11; Length 641;
 Best Local Similarity 24.4%; Pred. No. 6e-26;
 Matches 165; Conservative 59; Mismatches 190; Indels 263; Gaps 24;

```

QY 7 PEGALGYREFRHSDDVIGNINELRLR-GILTDVTLVGGPRLAHKAVLIAGSGFFS 65
DB 11 PEGC--YTVQVSRHSTEMLNINQKRGKGRFCVDLLRVGDESFPHRAVLAACSYFFS 68
QY 66 IFRGRAGV-----VDV---LSLPGP-----EARGFAPLLDMYTSRL 103
DB 69 VPSAQIGDGGADGGADVGGAAGAGGAGSRELEMTISSKVGIDLDPAVYSRIIV 128
QY 104 SPATPAVLAATVYLOMEHVQACHRFIOASYEPLGISRLPRLAEPRPTAP-----P 157
DB 129 RLSEFPELMTAKFLMRSVIEICQVIKOS-----NVQILVPPADIMLFRP 177
QY 158 PGSPRSSEGHPRPTESRSCSGCPSPBPAPKCMKKYKIVLNSQASQAGSLVGS 217
DB 178 PGT--SDLGF-----LDMTNGAMAMANSNGING 204
QY 218 SGQPCQARLPSCDEASSSSSSSSSSSECP--IPGQRLSPATAVQFKGAPASTY 276
DB 205 SMOPEEAAATGATAGASLPLVLPVDRLEPVAGP---LSFQLLTSPPNVASSAP- 260
QY 277 LITSQ-----AODTSGSPS--ERARPLP---GSEFFSQCNC--EAVAGCS- 315
DB 261 -LTSKRGGRPRKANLDMFSGPGLREAGILPGLCGKVFIDANRLQHEQHGVTS 319
QY 316 --GLDSLVP-----GDEKPYK-----COLCRSSFRYKGNLASHRTVHTG 353
DB 320 QLGIDLPRLGENGLPISEDPDGRKSRTRKQVACEICGKIFRDVYHLNHNKLSHS 379
QY 354 EKPYHSGIGARFNRPANLKTSHRISGE--KPYKETGSRFVYQRSQPSGQKPAR 411
DB 380 EKPYSPVCGLRKRDMSYHRSHDGSGVKPYIQSCGKF----- 422
QY 412 GGVGKGFCSGORODLKSPSOVAHLRAHV-LIHTEKPYPCPTGTRPHLQTLKSHV 470
DB 423 -----SRPDLNGHIKQVHTSERPHKQTCNASTATDRLSHL 461
QY 471 RHH----- 473
DB 462 ACHEDKVPQVCGKYLAAYMADHLKHSSEGFNFCSI CNREGQCSHODLIESSDYGD 521
QY 474 -----TGEPRYHCDPGLAFRHSQRLRLRQK 501
DB 522 LSDASDLKTRPKOSANGSPSCDVAVPKNKMSDEKKYPCPECGSFFRSKSYLNKHIQ-- 579
QY 502 HGAATNTKYVHILGSP 518
DB 580 -----KVHVRALGSP 589

```

RESULT 10

Q9HBE1 PRELIMINARY; PRT; 687 AA.
 AC Q9HBE1
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Zinc finger sarcoma gene long C isoform.
 GN ZSG.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 ON NCBI_TaxID=9606;
 RX MEDLINE=20404868; Pubmed=10949935;
 RA Mestrangeo T., Modena P., Tornelli S., Bullrich F., Teetl A.,
 RA Mezzei A., Radice P., Azzerelli A., Pilotti S., Croce C.,
 RA Pierotti M., Sozzi G.;

RT "A novel zinc finger gene is fused to EMS in small round cell tumor."
 RL Oncogene 19:3799-3804(2000).
 DR EMBL; AF254085; AAG09034.1;
 DR HSSP; 005516; ICS3.
 DR InterPro; IPR000637; AT hook.
 DR InterPro; IPR000515; BpD_transp.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR000822; ZnF_C2H2.
 DR Pfam; PF02178; AT hook; 1.
 DR Pfam; PF00651; BTB; 1.
 DR Pfam; PF00096; zF-C2H2; 8.
 DR ProDom; PD000003; ZnF_C2H2; 1.
 DR SMART; SM00384; AT_hook; 2.
 DR SMART; SM00225; BTB; 1.
 DR SMART; SM00355; ZnF_C2H2; 8.
 DR PROSITE; PS00402; BpD_TRANS_INN_MEMBER; UNKNOWN_1.
 DR PROSITE; PS00354; HMG1_Y; 1.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 6.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 7.
 KW DNA-binding; Metal-binding; Zinc-finger.
 SQ SEQUENCE 687 AA; 74060 MW; 0F1C59A78C6D110A CRC64;

Query Match 16.0%; Score 447; DB 4; Length 687;
 Best Local Similarity 26.6%; Pred. No. 3.9e-25;
 Matches 158; Conservative 60; Mismatches 193; Indels 184; Gaps 25;

```

QY 7 PEGALGYREFRHSDDVIGNINELRLR-GILTDVTLVGGPRLAHKAVLIAGSGFFS 65
DB 11 PEGC--YTVQVSRHSTEMLNINQKRGKGRFCVDLLRVGDESFPHRAVLAACSYFFS 68
QY 66 IFRGRAGV-----VDV---LSLPGP-----EARGFAPLLDMYTSRL 103
DB 69 VPSAQIGDGGADGGADVGATGATAPGGAGSRELEMTISSKVGIDLDPAVYSRIIV 128
QY 104 SPATPAVLAATVYLOMEHVQACHRFIOASYEPLGISRLPRLAEPRPTAP-----P 157
DB 129 RLSEFPELMTAKFLMRSVIEICQVIKOS-----NVQILVPPADIMLFRP 177
QY 158 PGSPRSSEGHPRPTESRSC--SGSPSPASBPDPKACNMKKYKIVLNSQASQAGSLV 214
DB 178 PGT--SDLGFPLDMTNGAALANSNGIAGSQPEEA-----ARAAGATA 221
QY 215 ERSQGQPCQA-RLSCDEASS--SSSSSSSEEGPIPGQRLSPATAVQFKGSA 270
DB 222 GQASLPLVPGVRLRMVAGPLSPQLTSPFVSAASAPRLTKRGGRPRKAN----- 274
QY 271 PASTPYLLTQAQDTSGSPS--ERARPLP---GSEFFSQCNC--EAVAGCS--GLD 318
DB 275 -----LDSM---FGSPGLREAGILPGLCGKVFIDANRLQHEQHGVTSIQLGVT 324
QY 319 SLVP-----GDEKPYK-----COLCRSSFRYKGNLASHRTVHTGEKPH 358
DB 325 DLPPRLGENGLPISEDPDGRKSRTRKQVACEICGKIFRDVYHLNHNKLSHSSEKPS 384
QY 359 CSICGARFNRPANLKTSHRISGE--KPYKETGSRFVYQRSQPSGQKPARGVQ 416
DB 385 CVCGGLRFRKRDMSYHRSHDGSGVKPYIQSCGKF----- 422
QY 417 KSGFCSGORODLKSPSOVAHLRAHV-LIHTEKPYPCPTGTRPHLQTLKSHVRIHTG 475
DB 423 -----SRPDLNGHIKQVHTSERPHKQTCNASTATDRLSHLACHED 466
QY 476 EKPYHCDPGLH-----FRHSQRLRLRQK 503
DB 467 KVP--QVCGKYLAAYMADHLKHSSEGFNFCSI CNRGFSASYLKVHVKTHHG 519

```

RESULT 11

Q9R161 PRELIMINARY; PRT; 703 AA.
 ID Q9R161
 AC Q9R161
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT	01-MAY-2000	(TREMBLrel. 13)	Last sequence update)
DT	01-JUN-2002	(TREMBLrel. 21)	Last annotation update)
DE	Zinc_finger protein ZFP233.		
GN	ZFP233.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxId=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Hauser L., York M., Shannon M., Stubbs L.;		
RT	"Differential expansion of homologous zinc-finger gene families in		
RL	human chromosome 19q13.2 and mouse chromosome 7."		
RL	Submitted (JUL-1999) to the EMBL/GenBank/DBD databases.		
CC	-1 SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).		
EMBL	AF167321; AACD45930.1; -.		
DR	HSSP; P25490; 1ZNM.		
DR	MGD; MGI:1929117; ZFP235.		
DR	InterPro; IPR001909; KRAB.		
DR	InterPro; IPR000822; Znf_C2H2.		
DR	Pfam; PF01352; KRAB; 1.		
DR	Pfam; PF00096; Zf-C2H2; 15.		
DR	PRINTS; PR00048; ZINC_FINGER.		
DR	ProDom; PD000003; Znf_C2H2; 12.		
DR	SMART; SMO0349; KRAB; 1.		
DR	SMART; SMO0355; Znf_C2H2; 15.		
DR	PROSITE; PS50805; KRAB; 1.		
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.		
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 15.		
DR	DNA-binding; Metal-binding; Nuclear protein; zinc-finger.		
QO	SEQUENCE 703 AA; 79398 MW; LAIDA000DID07386 CRC64;		

Query Match	15.9%;	Score 442.5;	DB 11;	Length 703;
Best Local Similarity	28.5%;	Pred. No. 8	6e-25;	
Matches 117;	Conservative 52;	Mismatches 142;	Indels 99;	Gaps 14;

```

QY      201 VLNSQASQASLVLGER-----SSGQPCFQARLP-----SGCEASSSS 237
           : : : : :
Db      185 ILNCISHQDNMLIHKDKDKPSSD-CDQYIFPMALCTLSPVYTBKAYQCTGTGDAITDS 243
           : : : : :
QY      238 SSSSSSS--SEEGPIPGQSRSLSPMAVQ-----FKGC-APASTPYLLTSQ 281
           : : : : :
Db      244 PSQGTQCVLLGNKSPVHGTYEDTSTSSVIOQSVPHKRRKYWCQECGKAFQSQSNLQTHQ 303
           : : : : :
QY      282 AQDTSSGP-----SERAPLP--GSEFFSCQNCCEAVAGSSSLDLV--P 322
           : : : : :
Db      304 RVHTGEKPYLTCPECGKSFNQSSHLVAMLEPIHNGEKPYCCDNCG--KGFSSRTDILNHCRV 361
           : : : : :
QY      323 GDEKRPYCCOLGRSSFRVYKGNLASHHTVYTGEEKPHYCIGCARFNPRLNLTHSRIRHSGE 362
           : : : : :
Db      362 HTGEEKPYKVCVCGKGTQSRSHLOAHERITGEEKPYKCGDCCGRKSCSSNLTHQRVHTEE 421
           : : : : :
QY      383 KPYKCEYCGSRF-----VQYRSQPPGFGQKPARGGVCGKGFSCSSQCRDILKSP 431
           : : : : :
Db      422 KPYKCECGKGRFLSLFNLSHQRVHTGEK--YCCECGKGFSSASSFSQSHQRVHTGEK 478
           : : : : :
QY      432 P-----SQVAHLRAHVLITNGEKDYPCCPTCGTRFRLQLTKSHVRITNGEKDYPHC 481
           : : : : :
Db      479 PFCCSVCGKGFSSSYFOAHQRVHTEKEPYRCDVCGKRFNLSLNLNHHQRVHTEGERPYKC 538
           : : : : :
QY      482 DPCGLHFRHKSQRLH-----LQKKGATNTKKNYHNLIG 517
           : : : : :
Db      539 EECCGKGSQASNLQAHQSVHTGEKPEKPCNAQCKRFSQASHLOAHRVHTG 588
           : : : : :

```

RESULT 12			
Q9HBE2	Q9HBE2	PRELIMINARY;	PRT; 537 AA.
AC	Q9HBE2;		
DT	01-MAR-2001	(T-EMBLrel. 16, Created)	
DT	01-MAR-2001	(T-EMBLrel. 16, Last sequence update)	
DT	01-JUN-2002	(T-EMBLrel. 21, Last annotation update)	
DE	Zinc finger sarcoma gene long B isoform.		

```

CN ZSG.
OS Homo sapiens (Human).
CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20404868; PubMed=10949935;
RA Matrangelo T., Modena P., Tornisella S., Bullrich F., Tassi A.,
RA Mezzelani A., Radice P., Azarelli A., Pilotti S., Croce C.,
RA Pierotti M., Sorzi G.,
RT "A novel zinc finger gene is fused to EWS in small round cell tumor.;"
RL Oncogene 19:3799-3804(2000).
DR EMBL; AF254084; AACG9033.1; -.
DR HSSP; Q05516; ICS3.
DR InterPro; IPR000637; AT hook.
DR InterPro; IPR000515; BPD_transp.
DR InterPro; IPR000210; BTB_POZ.
DR InterPro; IPR000822; Znf_C2H2.
DR Pfam; PF02178; AT_hook.1.
DR Pfam; PF00651; BTB.1.
DR Pfam; PF00096; zf_C2H2.6.
DR ProDom; PD000003; Znf_C2H2.1.
DR SMART; SM00384; AT_hook.2.
DR SMART; SM00225; BTB.1.
DR SMART; SM00355; Znf_C2H2.6.
DR PROSITE; PS00402; BPD_TRANSP_INN_MEMBER; UNKNOWN_1.
DR PROSITE; PS50097; BTB.1.
DR PROSITE; PS00354; HMGI_Y.1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1.4.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2.5.
KW DNA-binding; Metal-binding; Zinc-finger.
QO SEQUENCE 537 AA; 57995 MW; 950792413BFLFSD CRC64;

```

Query Match	15.8%	Score 442;	DB 4;	Length 537;
Best Local Similarity	27.3%	Pred. No. 6.9e-25;		
Matches 156; Conservative	61;	Mismatches 197;	Indels 158;	Gaps 25;

Qy	7	PEALQYVEEFGNRRSSDVLGNINLETRL-GULTVTLVLVGQPLRAKAVILCSGFYS	65
Db	11	PSCG--YTTQVSRHSTEMLHNINQGRKNXGFCVDLLRVGDSESPARAVIACSEYFEES	68
Qy	66	IFRGRRGVG-----VDV--LSLPGP-----EANGFPLDPMYTSRL	103
Db	69	VFAAQGDGDAADGADGADGATAPAGGAGSGRELEMHNTISSKVFDDILDFAYTSRIIV	128
Qy	104	SPATPAVLAATAATYLOMEHVVOACHRFIOASVEYLGISLRLEAPPTPTAP-----P	157
Db	129	RLESFELMTAAKFLIERSVIEICQEVIKG-----NQGLVPRABDMLFRP	177
Qy	158	PGSPRRSECHDPPTESNSC--SGPPSPASPDPKACNMKKYKIVLNSQASQAGSLVG	214
Db	178	PGT--SDLCFPLDMTNGALAAANSNGISGMQPESEA-----ARAGKALIA	221
Qy	215	ERSSGQPCQQA-RLPSGDEASS--SSSSSSSSEEGPIPGPOSRLSPTAATVQFKCGA	270
Db	222	GQASLPELVPGVRLPEWVAGPLSPOLLTTPFSVASSAPPLTGKRGGRPRKAN-----	274
Qy	271	PASTPYLLTSAODTSGSPS--ERARPLP---GSEFFSPQNC--EAVAGCS--GLD	318
Db	275	-----LIDSM--FGSFGGLREAGILPCGLCKVFEDANRLRQHEAQCIVTSLQGLYI	324
Qy	319	SLVP-----GDEKPKYK-----CQLCSSPPYKKNLASHHTVHTGKPYH	356
Db	325	DLPEPPLGENGPISEBDPGPKRSRTKQYACELCGKIFEDVYHLNRHLKLSHGKPYIS	384
Qy	359	CSICGARFNRPAWLKTHSRIRHSGE--KKYKCECTGSGRFVYORSQPSGFGKPARGVQ	416
Db	385	CPVCGARFKRKDRMSVYHRSHGSGVGRYICQSGCKG-----	422
Qy	417	KGGFGCSQODLKSPSOVAHLRAHV-LIHTGKPYEPPTCGTRFRHLQTLKLSHVRHTHG	475
Db	423	-----SRDHLNGIHKQVHSEBPHKQCTUNASFATRDRLRSIACHED	466

QY 476 EKYPHCDPCGLHFRHKSQRLHLRKHGATN 507
 DB 467 KVP--CQVCGKTLR-AAVADHLKKHSESPN 495

RESULT 13

Q9PVB8 PRELIMINARY; PRT; 470 AA.
 ID Q9PVB8;
 AC Q9PVB8;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Champion.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 OC Xenopodinae; Xenopus.
 OX NCBI_TaxID=8335;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21255550; PubMed=11357190;
 RT "A novel POZ/zinc finger protein, champion, interferes with
 RT gastrulation movements in Xenopus.";
 RL Dev. Dyn. 221:14-25(2001).
 DR EMBL; AB023074; BAA87059.1; -.
 DR HSP; P08047; ISP2.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00651; BTB; 1.
 DR Pfam; PF00096; zf-C2H2; 4.
 DR SMART; SM00225; BTB; 1.
 DR SMART; SM00355; Znf_C2H2; 4.
 DR PROSITE; PS50097; BTB; 1.
 DR PROSITE; PS50028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
 DR DNA-binding; zinc-finger.
 KW DNA-binding; zinc-finger.
 SQ SEQUENCE 470 AA; 52615 MW; 599121F68B2295FD CRC64;

Query Match 15.8%; Score 441; DB 13; Length 470;
 Best Local Similarity 26.6%; Pred. No. 7e-25;
 Matches 138; Conservative 57; Mismatches 184; Indels 140; Gaps 13;

QY 17 FTRHSDVIGNINELRLGILTDVTLVGQPLRAKAVLIACSGFFYSIFRBAVG-- 74
 DB 13 FPHSSDLSSLSNEQSHSGVLCITTKTGLEYPTRHRAVLAACSDYFRKFTGMPGRKC 72
 QY 75 ----VDVLSLPGPEARGAPLLDPMYTSRLSPATAVAVLAATYLMHVVQCHR 129
 DB 73 PDVCGDPLK---PALG--ALDDPAYTATLTISGNMMDVRAARLEIPCVVACVD 126
 QY 130 FIOAS--YEPGLISLRPLEAPPTPTAPPSPRSSEGHDPPTESRSCSQGPSPASP 187
 DB 127 ILQCNHREMGDADLECF-----LBARQVLSYMGENTATPPRAESP 173
 QY 188 DPKACMWKXYKTVLMSQASQAGSLVGRSSGQPCQARLPSGDEASSSSSSSSSSSE 247
 DB 174 PPHPHN-----LPVPPKSVQIIPRRKKFQLQVNPFRRNQNLRLGADDSLERDASHA 227
 QY 248 GPLPGQSLSPATATVQFKG---APASTPYLLTSQADTSG---SPSERARPLPS 299
 DB 228 GSPPNPISLGYEYVADNGIGHTIFVPPSPREIISD--HETSDMGFHNPDLENV--- 283
 QY 300 EPPSCNCEAVAGCCSGGLSLVPGDDK-----PYICQLCRSSFRYKGLASHRTV 350
 DB 284 ----SAGLDV-----SDKLVKRRSOLPQPCVCHKIIGHAGRLPRHMT 324
 QY 351 HTEKPYHOSIGARFRPRLNLTSHRISGEKPYCETGSRFVQVRSPPGPGFGKXA 410
 DB 325 HTEKPPVCCVCTTRNDKLIHMKHGERPYCEHSARFL----- 369
 QY 411 RGVGKGKGFCSQRODLKSPSOVAHLRAHVLHTEKPYPCPTGTRFRLQLTKSHV 470

DB 370 -----HSDYLNKM 378
 QY 471 RHTEKPYHCDPCGLHFRHKSQRLHLRKHGATNTK 509
 DB 379 HLHTGRPECSLCHKAFKEDHLORHMKGQNCLEVRT 417

RESULT 14
 Q9QZ48 PRELIMINARY; PRT; 569 AA.
 ID Q9QZ48;
 AC Q9QZ48;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Osteoclast-derived zinc finger (OCZF) protein.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SPRAGUE-DAWLEY;
 RX MEDLINE=99408739; PubMed=10477728;
 RT "Osteoclast-derived zinc finger (OCZF) protein with POZ domain, a
 RT possible transcriptional repressor, is involved in
 RT osteoclastogenesis.";
 RL Blood 94:1987-1997(1999).
 DR EMBL; D88450; BAA6393.1; -.
 DR HSP; P08046; IAH1.
 DR InterPro; IPR000210; BTB_POZ.
 DR InterPro; IPR000822; Znf_C2H2.
 DR Pfam; PF00651; BTB; 1.
 DR Pfam; PF00096; zf-C2H2; 4.
 DR SMART; SM00225; BTB; 1.
 DR SMART; SM00355; Znf_C2H2; 3.
 DR PROSITE; PS50097; BTB; 1.
 DR PROSITE; PS50028; ZINC_FINGER_C2H2_1; UNKNOWN_3.
 DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
 DR DNA-binding; zinc-finger.
 KW DNA-binding; zinc-finger.
 SQ SEQUENCE 569 AA; 60543 MW; C111AB303B0D0F0D CRC64;

Query Match 15.8%; Score 440.5; DB 11; Length 569;
 Best Local Similarity 25.2%; Pred. No. 9.5e-25;
 Matches 146; Conservative 57; Mismatches 186; Indels 191; Gaps 17;

QY 8 EGALGYVREFTRHSSDVIGNINELRLGILTDVTLVGQPLRAKAVLIACSGFFYSIF 67
 DB 6 DGPIGI--PFPDSSDILSGINEGRTOGLCDVILVGRREPPTSRVLAACSYFKLF 63
 QY 68 RRAVG-----VDVLSLPGPEARGAPLLDPMYTSRLSPATAVAVLAATYLMH 121
 DB 64 TSGAVALDOONVEIDFVS-----AEALTALMDFYTVTLVSTNVGDILSAALLLEIP 117
 QY 122 HVQACHRFIOASYPLGISLRPLEAPPTPTAPPSPRSSEGHDPPTESRSCSQGP 181
 DB 118 AVSRVCTDLE-----RQILAADDVGAQPDGA-----CPTDQRULRLRK 158
 QY 182 -----PSPASDPNACWKY-----KYIVNSQASQAGSLVGRSSGQ 221
 DB 159 EYLFFRSNPNMSLP-PTAFQWPGFSAPDDDLDAATKEAVAAVAALVAGDCGLPFGP 217
 QY 222 CPOALPSGDEASSSSSSSSSSSEGLPG-----PQRLSTATAVQFKG--APASTPY 276
 DB 218 PPADRPPTGDEGGSTGLWPERDEDAAPGGLFPPT--APPATYNGHYRAGAST-- 273
 QY 277 ILTSQADPTSGSPSRARPLPG--SEFFSCQNCCEAVAGCCSGGLSLV----- 321
 DB 274 -----GEEBAVALSE--AAPBQDSGFLSGAAGDGDAADVGLAATLLOQMSSVGR 327
 QY 322 PGDEDKP-----YKCOLCRSSFRY 340

DB 328 AGSDSESRPDDKGVMDYLLKYPGAEHGDVYPAWSQKGEKKIRAKAFQKPICEKVIQ 387
QY 341 KGNLASRVTHTGEKPYHCSICGARFNRPAHLKTHSRIRHSGEKPYKCEGSHFVQVRG 400
DB 388 AGULPHRIHTHTEKPYECNICVRFTRQDLKYHNRKHTGEKPYLCQCGAFA----- 442
QY 401 PPSGFOGKPARGVGQKGFSSQRODLKSPSQVAHLRAHVLHTHGEKPYPCPTGTRP 460
DB 443 ----- 442
QY 461 RHLQTLKSHVRIHTGEKPYHCDPCGLHFRHKSQLRLHLRQ 500
DB 443 -HNYDLKNHMRVHTGLRPYQDCCKTFVNSDHLHRLKK 481
RESULT 15
ID 09H117 PRELIMINARY; PRT; 545 AA.
AC 09H117;
DT 01-MAR-2001 (TEMBLrel. 16, Created)
DT 01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE D0322G13.2.1 (zinc finger protein FLJ21794, isoform 1).
GN D0322G13.2.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Lovell J.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL096677; CAC17422.1; -
DR HSSP: P07248; 1ARF.
DR InterPro: IPR000210; BTB_POZ.
DR InterPro: IPR000822; ZNF_C2H2.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF00096; zf-C2H2; 8.
DR ProDom: PD000003; Znf_C2H2; 1.
DR SMART: SM00325; BTB; 1.
DR SMART: SM00355; ZNF_C2H2; 8.
DR PROSITE: PS50097; BTB; 1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 7.
DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 8.
DR DNA-binding; Metal-binding; Zinc-finger.
SQ SEQUENCE 545 AA; 61982 MW; AAA9820E319C361E CRC64;
Query Match 15.8%; Score 439.5; DB 4; Length 545;
Best Local Similarity 25.2%; Pred. NO. 1.1e-24;
Matches 149; Conservative 61; Mismatches 200; Indels 181; Gaps 19;
QY 23 DVLGNINELRLRGLDVTLLVGOPLR---AHKAVLIACSGPFYSIFRGRAGV--GVDV 77
DB 16 NLHHEHMEHLRLHGLCLCVTVSVYQGRKDFMAHKAVALATSKFKEVFLNEKSVDTGTR 75
QY 78 LSLPGGEAAGAPALDPMYTSRLSPATAPAVLAATYIQMEHVQACHRFIOASYEP 137
DB 76 NVYLINEVQVADFPASPLEFYTAQVIEDRYQRMLEVAEKLKCLDLSCTFQLKQMLSS 135
QY 138 LGISLRLP---EAEPPT---PPTAPPSSFRSEGHDPPTES-----RSCSGQPPSP 184
DB 136 VLLLELQNFSESQEEVSSGSQVSAAPRASVATDGPHPGLTDSLDPGERASNGMSSD 195
QY 185 ASDPDKACNMKKYKTYI-----LNSQASQAGSLV 213
DB 196 LPPKSKDKLDKKKEVKKPPYKIRASGRLAGRKVVEIPKKKYTRRLREQQKTAEGDV 255
QY 214 GERSSGQPCQPARLPSS-----DEASSSSSSSSSSSSSECP1PGP 253
DB 256 GD---YRCFQDQSPRVGTGEMEQVSKNECGQAGALEBELSKKAGPBEHEEEDDEGE 311
QY 254 QSRLSPTAATVQFKC-----GAPASTPYLLTSAQDTSSGP--- 289

DB 312 KKKSN-----FKSICEKAFLEYKSPFKSHKRGVATEVVYRC-----DTCGQTPAN 359
QY 290 ---SEBARPLPGSE--FPSQONC-----EAVAGSSGLDVLVPGDEDEKPYKQL 333
DB 360 RGNLKSQDRAVHSSERHFPCELCGKKFKKQDVKNHVLQVHEG-----GSE--RHRGQ 411
QY 334 CRSSFRYKGNLASRVTHTGEKPYHCSICGARFNRPAHLKTHSRIRHSGEKPYKCEGSH 393
DB 412 CGKGLSKTALRLHRTHTGDRPYGCTEGGARFSQPSALKTMRHTHGEKPYFVCDCEGAR 471
QY 394 FVQVNSQPPSGQKPARGVGQKGFSSQRODLKSPSQVAHLRAHVL-----HTGE 448
DB 472 FTQ-----NHMLLYHKKCHTGE 488
QY 449 KPYPCTGTRPRHLQTLKSHVRIHTGEKPYHCDPCGLHFRHKSQLRLHLR 499
DB 489 RFPKCECTCKSPASKKEYLKNHNRHTGSKFPKCEVCFTFAORNSLYGHK 539

Search completed: July 14, 2003, 18:20:56
Job time : 27.0314 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 14, 2003, 18:02:16 ; Search time 71.3175 Seconds
(without alignments)
3843.329 Million cell updates/sec

Title: US-09-815-379-10

Perfect score: 10730
Sequence: 1 FCLQGRVLRNQGHPST.....MIVKKRYSTRSASSQSSR 2057

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: A.Geneseq.101002.*
2: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1980.DAT.*
3: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1981.DAT.*
4: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1982.DAT.*
5: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1983.DAT.*
6: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1984.DAT.*
7: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1985.DAT.*
8: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1986.DAT.*
9: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1987.DAT.*
10: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1988.DAT.*
11: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1989.DAT.*
12: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1990.DAT.*
13: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1991.DAT.*
14: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1992.DAT.*
15: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1993.DAT.*
16: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1994.DAT.*
17: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1995.DAT.*
18: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1996.DAT.*
19: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1997.DAT.*
20: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1998.DAT.*
21: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA1999.DAT.*
22: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA2000.DAT.*
23: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA2001.DAT.*
24: /SIDS2/gcgdata/geneeq/geneeqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	10730	100.0	2057	AAE11890	Angiogenesis assoc
2	10626	99.0	2058	ABE97219	Novel human protei
3	10247	95.5	2048	AAE11891	Angiogenesis assoc
4	7166.5	66.8	2408	ABG10631	Novel human diagno
5	6775.5	63.1	1515	ABG10624	Novel human diagno
6	5385	50.2	1031	AAU17082	Novel signal trans
7	5166	48.1	984	AAW93762	Human polypeptide
8	5146.5	48.0	1442	ABG25601	Novel human diagno
9	4450	41.5	869	ABG42020	Human ORFX ORF1784
10	3238.5	30.2	713	ABG10632	Novel human diagno

11	2032	18.9	394	22	AAE93020
12	1947	18.1	402	22	AAU23129
13	1843.5	17.2	2167	22	ABE60369
14	1789	16.7	348	22	AAU17508
15	1705	15.9	2129	22	ABE62828
16	1495	13.9	2424	22	ABE58924
17	1492	13.9	2548	20	AAV05781
18	1476	13.8	1544	21	AAE48638
19	1466.5	13.7	1493	21	AAE48639
20	1450.5	13.5	1556	21	AAE30438
21	1444.5	13.5	1483	21	AAE48640
22	1441	13.4	1505	21	AAE30439
23	1434.5	13.4	697	17	AAW00603
24	1426	13.3	631	17	AAW00604
25	1419	13.2	1495	21	AAE30440
26	1372.5	12.8	1792	22	ABE71108
27	1352	12.6	1839	21	AAE94291
28	1344.5	12.5	301	22	ABE25592
29	1322.5	12.3	1972	17	AAW00024
30	1313	12.2	2056	22	ABE59344
31	1308	12.2	1453	22	AAE39213
32	1306	12.2	1469	22	AAE39214
33	1305.5	12.2	1754	21	AAE52410
34	1305.5	12.2	1769	21	AAE52409
35	1305.5	12.2	1804	21	AAE52408
36	1303	12.1	1988	22	AAE40999
37	1303	12.1	1988	22	AAE40999
38	1297.5	12.1	1960	22	AAE78854
39	1297.5	12.1	1963	22	AAE78838
40	1274	11.9	1939	23	ABE77096
41	1269	11.8	1948	22	ABE21233
42	1268	11.8	1879	22	AAE25750
43	1262.5	11.8	1120	21	AAE94292
44	1261	11.8	2067	22	ABE71125
45	1249.5	11.6	1857	23	AAU84350

ALIGNMENTS

RESULT 1	
AAE11890	
ID	AAE11890 standard; Protein; 2057 AA.
AC	AAE11890;
DT	18-DEC-2001 (first entry)
DE	Angiogenesis associated human myosin XI (hmxi) protein.
XX	Angiogenesis associated protein; AAP; cytosolic; cardiac; gene therapy;
KW	ophthalmological; valvular; myocardial infarction; macular degeneration;
KW	diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine;
KW	rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation;
KW	cancer; therapeutic; diagnostic; human; myosin XI; MXI.
OS	Homo sapiens.
XX	
PN	WO200170808-A2.
XX	
PD	27-SEP-2001.
XX	
PF	22-MAR-2001; 2001WO-US09609.
XX	
FR	22-MAR-2000; 2000US-191134P.
XX	
PA	(CURA-) CURAGEN CORP.
PA	(GUTH) GENENTECH INC.
XX	
PI	Rastelli LK, Gerritsen M;
XX	
XX	WPI; 2001-602775/68.
DR	N-PSDB; AAD19119.

Human protein sequ
Novel human enzyme
Arabidopsis melano
Novel signal trans
Drosophila melano
Drosophila melano
Human myosin IXa.
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human non-conventi
Human myosin VII
Arabidopsis thalia
Drosophila melano
Caenorhabditis ele
Novel human diagno
Smooth muscle myos
Human polypeptide
Human polypeptide
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Human polypeptide
Human polypeptide
Human protein SEQ
Human alpha-myosin
Novel human diagno
Human protein sequ
Helianthus annuus
Arabidopsis melano
Protein MYH1 diff

XX Novel angiogenesis associated polypeptides and polynucleotides encoding
PT the polypeptides, useful for modulating angiogenesis and for treating
PT tumors and cancers -
XX
PS Claim 1, Page 31-38, 159pp; English.
XX
CC The invention relates to angiogenesis associated proteins (AAP) and their
CC corresponding cDNA molecules, which are useful for modulating
CC angiogenesis. AAP proteins and nucleic acids are useful for promoting
CC wound healing, for example after organ transplantation, and in the
CC treatment of tumors, myocardial infarction, cancers, diabetic
CC retinopathy, muscular degeneration, psoriasis and rheumatoid arthritis.
CC AAP proteins and DNA's are useful in potential prophylactic and
CC therapeutic applications implicated in a variety of disorders including
CC those related to angiogenesis, and also in diagnostic applications.
CC AAP cDNA is also useful in gene therapy. The invention also relates to
CC a method for screening a tissue sample for tumorigenic potential. AAP
CC proteins are used to screen drugs or compounds that modulate AAP activity
CC or expression as well as treating disorders characterized by insufficient
CC or excessive production of AAP or production of AAP forms that have
CC decreased or aberrant activity compared to the wild type protein, or
CC modulate biological function that involve AAP. The present sequence
CC is human myosin XI (hmx1) protein which is an angiogenesis associated
CC protein (AAP) of the invention. Hmx1 is upregulated in the in vitro model
CC of angiogenesis and is likely to be involved in transducing signals from
CC angiogenic factors, perhaps modulating the cytoskeleton.
CC
XX
SQ Sequence 2057 AA:
Query Match 100.0%; Score 10730; DB 22; Length 2057;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2057; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 FCTGTRVWLRKNGQHPPTVNSCAGIYVTRDYGOVFTYKOSTTHOKVTMHPNNE 60
DB 1 FCTGTRVWLRKNGQHPPTVNSCAGIYVTRDYGOVFTYKOSTTHOKVTMHPNNE 60
QY 61 GVDNMASTLTLHGSGIMYNLFGYKRNQWITYGSLASVNPQIAGLEPAPMEOYSR 120
DB 61 GVDNMASTLTLHGSGIMYNLFGYKRNQWITYGSLASVNPQIAGLEPAPMEOYSR 120
QY 121 RHIGELPRHIFAIANECYRCLMKRHDNOCILIGESGAGTSTKILKLTSLYSQOSLE 180
DB 121 RHIGELPRHIFAIANECYRCLMKRHDNOCILIGESGAGTSTKILKLTSLYSQOSLE 180
QY 121 RHIGELPRHIFAIANECYRCLMKRHDNOCILIGESGAGTSTKILKLTSLYSQOSLE 180
DB 121 RHIGELPRHIFAIANECYRCLMKRHDNOCILIGESGAGTSTKILKLTSLYSQOSLE 180
QY 181 LSLKERTSCYERAILLESSPIWEAFGNAKTYVNNSSRFKGVOLNCOKNGIGGRILVC 240
DB 181 LSLKERTSCYERAILLESSPIWEAFGNAKTYVNNSSRFKGVOLNCOKNGIGGRILVC 240
QY 241 ILSSQRVVRQNGERNYHIFVALLAGLEHEEREPEFLSTPENYHYLNOSGVEDTITSD 300
DB 241 ILSSQRVVRQNGERNYHIFVALLAGLEHEEREPEFLSTPENYHYLNOSGVEDTITSD 300
QY 301 QESFREYITMDWQSKSEYREVSRLLAGILHGNIEFTTAGAQSFTALGRSABEL 360
DB 301 QESFREYITMDWQSKSEYREVSRLLAGILHGNIEFTTAGAQSFTALGRSABEL 360
QY 301 QESFREYITMDWQSKSEYREVSRLLAGILHGNIEFTTAGAQSFTALGRSABEL 360
DB 301 QESFREYITMDWQSKSEYREVSRLLAGILHGNIEFTTAGAQSFTALGRSABEL 360
QY 361 GLDPTOLTDLTORSMFLRGEELITPLNQAVDSRSLAMALYACCFEVIKIKISRIK 420
DB 361 GLDPTOLTDLTORSMFLRGEELITPLNQAVDSRSLAMALYACCFEVIKIKISRIK 420
QY 421 GNEDFKSIGILDI FGENPEVNFHFOFNINANEKLOEYFNKHI FSLQLEYSREGIWE 480
DB 421 GNEDFKSIGILDI FGENPEVNFHFOFNINANEKLOEYFNKHI FSLQLEYSREGIWE 480
QY 481 DIWINDGECLDIIEKKGLGALAINESHPQATDSTLLEKLSQHANNHFFYKPVAVN 540
DB 481 DIWINDGECLDIIEKKGLGALAINESHPQATDSTLLEKLSQHANNHFFYKPVAVN 540
QY 541 NFGVKHAYAGEVOYDVNGIIEKNDTPRDDLLNLRSPDPIYDLFEHVSRRNOOTLKC 600
DB 541 NFGVKHAYAGEVOYDVNGIIEKNDTPRDDLLNLRSPDPIYDLFEHVSRRNOOTLKC 600

QY 601 GSKHRRPTVSSQFXYDSLHSLMATTSSNPFVRCIKPMQMPQPDQAVVLNQLRYSG 660
DB 601 GSKHRRPTVSSQFXYDSLHSLMATTSSNPFVRCIKPMQMPQPDQAVVLNQLRYSG 660
QY 661 METVIRIRKAGYAVRRPDDFYKRYKVLNRNLALBEDVRGKCTSLQLYDASNEMQLK 720
DB 661 METVIRIRKAGYAVRRPDDFYKRYKVLNRNLALBEDVRGKCTSLQLYDASNEMQLK 720
QY 721 TYVFLRESLEQLEKRESEVSHAMVIRAHVIGFLARKQYRVYCVIIOKNRPAFL 780
DB 721 TYVFLRESLEQLEKRESEVSHAMVIRAHVIGFLARKQYRVYCVIIOKNRPAFL 780
QY 781 RRRFLHLKXAAIVFOKQNGQIARRVYRQLAEKREBEKKKQREBEKKKREBEKKRER 840
DB 781 RRRFLHLKXAAIVFOKQNGQIARRVYRQLAEKREBEKKKQREBEKKKREBEKKRER 840
QY 841 ERREAEALRQOEERFRKQOELALQSQAEALTELEKQKNKQVEILRLKEIEDIQ 900
DB 841 ERREAEALRQOEERFRKQOELALQSQAEALTELEKQKNKQVEILRLKEIEDIQ 900
QY 901 RKKEQOELSLTASLOKQERDOELRLREBACRAOEFSLNPFDEIDECVRNTERSL 960
DB 901 RKKEQOELSLTASLOKQERDOELRLREBACRAOEFSLNPFDEIDECVRNTERSL 960
QY 961 SCGSEFSSSLASACEKENFNSQYPPEEYDEGFEADDDAFKDSPNSEHSHDQRTS 1020
DB 961 SCGSEFSSSLASACEKENFNSQYPPEEYDEGFEADDDAFKDSPNSEHSHDQRTS 1020
QY 1021 GRTGDDSEEDPDMNDVTVPTSPADSTVLLAPVQSGSLHNSSGSSTYCMQONMGD 1080
DB 1021 GRTGDDSEEDPDMNDVTVPTSPADSTVLLAPVQSGSLHNSSGSSTYCMQONMGD 1080
QY 1081 LPSPDGDYDDODDEDGALITSGSSVTFSNSGQSMSPDYRCSVGTYNSSGAYRESSGA 1140
DB 1081 LPSPDGDYDDODDEDGALITSGSSVTFSNSGQSMSPDYRCSVGTYNSSGAYRESSGA 1140
QY 1141 QGSFEDSEEDPDRFDTDELSYRDSVYSCTLPYFHSFLYMKGLMNSWRRKRCVLKD 1200
DB 1141 QGSFEDSEEDPDRFDTDELSYRDSVYSCTLPYFHSFLYMKGLMNSWRRKRCVLKD 1200
QY 1201 EFTFLFRSQOELKQGLHKKGGSSSTLSRRWKRWPFLRQSLMYFENDSEELKGV 1260
DB 1201 EFTFLFRSQOELKQGLHKKGGSSSTLSRRWKRWPFLRQSLMYFENDSEELKGV 1260
QY 1261 EYRTAKEIIDNTTKNGIDIIMADRPHLIAESPEDASQWFSLVQVHASTDOEIOEMHD 1320
DB 1261 EYRTAKEIIDNTTKNGIDIIMADRPHLIAESPEDASQWFSLVQVHASTDOEIOEMHD 1320
QY 1321 EQANPQNAVGLDVGILDSVCASDSPDRPNSFVITIANRVLHCNADTEBEMHWITLLOR 1380
DB 1321 EQANPQNAVGLDVGILDSVCASDSPDRPNSFVITIANRVLHCNADTEBEMHWITLLOR 1380
QY 1381 SKGDRRVGEQEFIVGMLHKEVNSPKMSLSLKKRWPLTNLSLDYKSSSKMLKIGT 1440
DB 1381 SKGDRRVGEQEFIVGMLHKEVNSPKMSLSLKKRWPLTNLSLDYKSSSKMLKIGT 1440
QY 1441 LVINSLCSVVPDEKIFKETGYMNTVYGRKHCYRLYTKLNEATRMSVIONVDTKAP 1500
DB 1441 LVINSLCSVVPDEKIFKETGYMNTVYGRKHCYRLYTKLNEATRMSVIONVDTKAP 1500
QY 1501 IDPTQQLIODIKENCLNSDVEEQIYKGNPILRYTHHPLSPFLPYGDINLNLKDG 1560
DB 1501 IDPTQQLIODIKENCLNSDVEEQIYKGNPILRYTHHPLSPFLPYGDINLNLKDG 1560
QY 1561 YTTLODEAKIINSLQOLESMSDPIPIIOGIIQTHDRLPLDELYCQILKQTNVPHG 1620
DB 1561 YTTLODEAKIINSLQOLESMSDPIPIIOGIIQTHDRLPLDELYCQILKQTNVPHG 1620
QY 1621 SVGNLYSQOILTCLCTFLPSGILKYLKFLHKLIREOPGTEMEKVALFTYESLKTGC 1680
DB 1621 SVGNLYSQOILTCLCTFLPSGILKYLKFLHKLIREOPGTEMEKVALFTYESLKTGC 1680

QY 1681 REFVPSRDEIEALLHROEMTSIVYCHGGSCKITINSHTIAGEVVEKLIIRGLAMEDSRNM 1740
DB 1681 REFVPSRDEIEALLHROEMTSIVYCHGGSCKITINSHTIAGEVVEKLIIRGLAMEDSRNM 1740
QY 1741 PALFENGHVDAKISRRTAVADVLAKEKLAATSEVGDLEFKRPFYLYGFLDTDVPKDS 1800
DB 1741 PALFENGHVDAKISRRTAVADVLAKEKLAATSEVGDLEFKRPFYLYGFLDTDVPKDS 1800
QY 1801 VEFAPMEQAHBAVIGHHHPABENQVLAALRLOGLCYTLHAALPPLSEVYSIQRLK 1860
DB 1801 VEFAPMEQAHBAVIGHHHPABENQVLAALRLOGLCYTLHAALPPLSEVYSIQRLK 1860
QY 1861 ARISSSTKTFTPCERLEKERTSFLEGTLRSSFRGTSVACVKEBEQMLDMWKEEVSAR 1920
DB 1861 ARISSSTKTFTPCERLEKERTSFLEGTLRSSFRGTSVACVKEBEQMLDMWKEEVSAR 1920
QY 1921 ASIIDKMRKFGQMNQOAMAKYMALIKEMPGYSTLPDVECKEGGFPOELMGVSADAVS 1980
DB 1921 ASIIDKMRKFGQMNQOAMAKYMALIKEMPGYSTLPDVECKEGGFPOELMGVSADAVS 1980
QY 1981 VKRGEGRPLEVFOYEHILSPGAPLANTYKIVDERELLETSEVVDVAKLMKAYISMTV 2040
DB 1981 VKRGEGRPLEVFOYEHILSPGAPLANTYKIVDERELLETSEVVDVAKLMKAYISMTV 2040
QY 2041 KRRYSTTRGASSQSSSR 2057
DB 2041 KRRYSTTRGASSQSSSR 2057

RESULT 2
ABB97219
ID ABB97219 standard; Protein; 2058 AA.
AC ABB97219;
XX

DT 27-JUN-2002 (first entry)
XX
DB Novel human protein SEQ ID NO: 487.
XX

KM Human; anti-tumour; vulnotherapy; anti-inflammatory; immunomodulator;
XX anti-infectivity; cerebroprotective; cytostatic; rheumatic; gene therapy;
KM neuroprotective; anti-parasitism; protein therapy; EST;
XX expressed sequence tag.
OS Homo sapiens.
XX

PN WO200222660-A2.
XX

PD 21-MAR-2002.
XX

PF 10-SEP-2001; 2001WO-US26015.
XX

PR 11-SEP-2000; 2000US-0659671.
XX

XX (HYSB-) HYSBQ INC.
XX

PI Tang YF, Liu C, Zhou P, Aducci V, Zhang J, Zhao QA, Ren F;
XX Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX

DR WPI; 2002-292408/33.
XX

DR N-PSDB; ABB32405.
XX

XX An isolated polynucleotide for treating diseases associated with its
XX encoded polypeptide such as cancer and multiple sclerosis -
XX

PS Example 2; SEQ ID NO 487; 509pp; English.
XX

CC The present invention provides the protein and coding sequences of 444
XX novel human proteins. These were isolated from expressed sequences tags
XX (ESTs). They can be used to stimulate cell growth, to regulate
XX haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
XX e.g. in burn treatment, to regulate the immune system e.g. to treat
XX multiple sclerosis, to regulate activin or inhibin e.g. to treat

CC infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC Parkinson's disease. The present sequence is a protein of the invention.
XX

SO Sequence 2058 AA;

Query Match 99.0%; Score 10626; DB 23; Length 2058;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 2043; Conservative 4; Mismatches 8; Indels 2; Gaps 2;

QY 1 FCLQSTRVMBNGOHFSTVNSCAEGIVRTDYGQYFTTKOSTITHTQKTAMPTNE 60
DB 1 FCLQSTRVMBNGOHFSTVNSCAEGIVRTDYGQYFTTKOSTITHTQKTAMPTNE 60
QY 4 FTEGTIRVMBNGOHFSTVNSCAEGIVRTDYGQYFTTKOSTITHTQKTAMPTNE 63
DB 4 FTEGTIRVMBNGOHFSTVNSCAEGIVRTDYGQYFTTKOSTITHTQKTAMPTNE 63
QY 61 GVDMASTLHAGGIMVNLFORVKNQIMVYIGSLASVNPYOPDAGLYPATMEQYSR 120
DB 61 GVDMASTLHAGGIMVNLFORVKNQIMVYIGSLASVNPYOPDAGLYPATMEQYSR 120
QY 64 GVDMASTLHAGGIMVNLFORVKNQIMVYIGSLASVNPYOPDAGLYPATMEQYSR 123
DB 64 GVDMASTLHAGGIMVNLFORVKNQIMVYIGSLASVNPYOPDAGLYPATMEQYSR 123
QY 121 RHLGELPHIPALANECYRCMLKHDNOCILIKSGAGKTESTLILKPLSVISQSLB 180
DB 121 RHLGELPHIPALANECYRCMLKHDNOCILIKSGAGKTESTLILKPLSVISQSLB 180
QY 124 RHLGELPHIPALANECYRCMLKHDNOCILIKSGAGKTESTLILKPLSVISQSLB 183
DB 124 RHLGELPHIPALANECYRCMLKHDNOCILIKSGAGKTESTLILKPLSVISQSLB 183
QY 181 LSLKETSCEBRALIESPIMEAFGNACTVNNNSRFGKTVOLNICOXNIOGGRIVDC 240
DB 181 LSLKETSCEBRALIESPIMEAFGNACTVNNNSRFGKTVOLNICOXNIOGGRIVDC 240
QY 184 LSLKETSCEBRALIESPIMEAFGNACTVNNNSRFGKTVOLNICOXNIOGGRIVDC 243
DB 184 LSLKETSCEBRALIESPIMEAFGNACTVNNNSRFGKTVOLNICOXNIOGGRIVDC 243
QY 241 ILSSQNRVVRONPGERNHIFVALAGLEHREBEFYLTPEYHYLNQSGVEDKTSID 300
DB 241 ILSSQNRVVRONPGERNHIFVALAGLEHREBEFYLTPEYHYLNQSGVEDKTSID 300
QY 244 LL-EKORVRONPGERNHIFVALAGLEHREBEFYLTPEYHYLNQSGVEDKTSID 302
DB 244 LL-EKORVRONPGERNHIFVALAGLEHREBEFYLTPEYHYLNQSGVEDKTSID 302
QY 301 QESFEVITAMDVNQFSKEEVEVSRLAGILHLGIEFITAGAQVSPKALAGSABL 360
DB 301 QESFEVITAMDVNQFSKEEVEVSRLAGILHLGIEFITAGAQVSPKALAGSABL 360
QY 303 QESFEVITAMDVNQFSKEEVEVSRLAGILHLGIEFITAGAQVSPKALAGSABL 362
DB 303 QESFEVITAMDVNQFSKEEVEVSRLAGILHLGIEFITAGAQVSPKALAGSABL 362
QY 361 GLDPTQLTALTONSMPLRGEIILTPLVQOAVDSRSLAMALYACCEBWIYKINSR 420
DB 361 GLDPTQLTALTONSMPLRGEIILTPLVQOAVDSRSLAMALYACCEBWIYKINSR 420
QY 363 GLDPTQLTALTONSMPLRGEIILTPLVQOAVDSRSLAMALYACCEBWIYKINSR 422
DB 363 GLDPTQLTALTONSMPLRGEIILTPLVQOAVDSRSLAMALYACCEBWIYKINSR 422
QY 421 GNEDEKTSIGIIDIRGEFENFEVNHFEQFINVANELOEYFNKHIPSLQLEYSRGLWE 480
DB 421 GNEDEKTSIGIIDIRGEFENFEVNHFEQFINVANELOEYFNKHIPSLQLEYSRGLWE 480
QY 423 GNEDEKTSIGIIDIRGEFENFEVNHFEQFINVANELOEYFNKHIPSLQLEYSRGLWE 482
DB 423 GNEDEKTSIGIIDIRGEFENFEVNHFEQFINVANELOEYFNKHIPSLQLEYSRGLWE 482
QY 481 DIDWIDNGECUDLEKKGILALINEBSHPQATDSTLEKLSQHANNHFFVXPRVAVN 540
DB 481 DIDWIDNGECUDLEKKGILALINEBSHPQATDSTLEKLSQHANNHFFVXPRVAVN 540
QY 483 DIDWIDNGECUDLEKKGILALINEBSHPQATDSTLEKLSQHANNHFFVXPRVAVN 542
DB 483 DIDWIDNGECUDLEKKGILALINEBSHPQATDSTLEKLSQHANNHFFVXPRVAVN 542
QY 541 NFGVHYAGEVQYDVRGILEKNDTFRDILNLRBSRFDIYDLFEHVSSRNNQDTLKC 600
DB 541 NFGVHYAGEVQYDVRGILEKNDTFRDILNLRBSRFDIYDLFEHVSSRNNQDTLKC 600
QY 543 NFGVHYAGEVQYDVRGILEKNDTFRDILNLRBSRFDIYDLFEHVSSRNNQDTLKC 602
DB 543 NFGVHYAGEVQYDVRGILEKNDTFRDILNLRBSRFDIYDLFEHVSSRNNQDTLKC 602
QY 601 GSKRRPTVSSQFK-DLSHSLMATLSSNPFVRCIKRMQMPQFOQAVVNLQRLSG 660
DB 601 GSKRRPTVSSQFK-DLSHSLMATLSSNPFVRCIKRMQMPQFOQAVVNLQRLSG 660
QY 603 GSKRRPTVSSQFK-DLSHSLMATLSSNPFVRCIKRMQMPQFOQAVVNLQRLSG 661
DB 603 GSKRRPTVSSQFK-DLSHSLMATLSSNPFVRCIKRMQMPQFOQAVVNLQRLSG 661
QY 661 MLETYRIKAGYAARRPQDFYKRYKVLMMRMLALPEDVGRGCTSLGLYDASNSSEQGLK 720
DB 661 MLETYRIKAGYAARRPQDFYKRYKVLMMRMLALPEDVGRGCTSLGLYDASNSSEQGLK 720
QY 662 MLETYRIKAGYAARRPQDFYKRYKVLMMRMLALPEDVGRGCTSLGLYDASNSSEQGLK 721
DB 662 MLETYRIKAGYAARRPQDFYKRYKVLMMRMLALPEDVGRGCTSLGLYDASNSSEQGLK 721
QY 721 TKVFLRSLEQLEKREBEVSHAAVIRAHVGLFARKQYRKVLVCVVIIOKNYRAFL 780
DB 721 TKVFLRSLEQLEKREBEVSHAAVIRAHVGLFARKQYRKVLVCVVIIOKNYRAFL 780
QY 722 TKVFLRSLEQLEKREBEVSHAAVIRAHVGLFARKQYRKVLVCVVIIOKNYRAFL 781
DB 722 TKVFLRSLEQLEKREBEVSHAAVIRAHVGLFARKQYRKVLVCVVIIOKNYRAFL 781
QY 761 RRRFLHLKKAIVPQKQIRARVYVQQLAEXRQEQKQEQKQEQKQEQKQEQKQEQK 840
DB 761 RRRFLHLKKAIVPQKQIRARVYVQQLAEXRQEQKQEQKQEQKQEQKQEQKQEQK 840
QY 762 RRRFLHLKKAIVPQKQIRARVYVQQLAEXRQEQKQEQKQEQKQEQKQEQKQEQK 841
DB 762 RRRFLHLKKAIVPQKQIRARVYVQQLAEXRQEQKQEQKQEQKQEQKQEQKQEQK 841
QY 841 ERREALELAQOEEBTRKQOELEALQSQKEALELTRELEKQKQKQKQKQKQKQKQ 900
DB 841 ERREALELAQOEEBTRKQOELEALQSQKEALELTRELEKQKQKQKQKQKQKQKQ 900
QY 842 ERREALELAQOEEBTRKQOELEALQSQKEALELTRELEKQKQKQKQKQKQKQKQ 901
DB 842 ERREALELAQOEEBTRKQOELEALQSQKEALELTRELEKQKQKQKQKQKQKQKQ 901
QY 901 RMKEQOELSTBASIQKQERDDELRLLEBACRAAOEPLBSJNPFDEIDECVNIERSL 960
DB 901 RMKEQOELSTBASIQKQERDDELRLLEBACRAAOEPLBSJNPFDEIDECVNIERSL 960

Db 902 RMKEOELSLTEASLOKLOERRDOELRLLEEACRAOEFLSLNDEIDECVRNTERSL 961
QY 961 SGGSEFSSSLAESAACEKPNFNFSPYPEEVEGGEFADDAFKDSPNPEHSGDQRTS 1020
Db 962 SVGSSEFSSSLAESAACEKPNFNFSPYPEEVEGGEFADDAFKDSPNPEHSGDQRTS 1021
QY 1021 G1RTSDSSSEEDPYMDTVPTSPADSTVLLAPSVODSGSLHNSSGSESTYCPONAGD 1080
Db 1022 G1RTSDSSSEEDPYMDTVPTSPADSTVLLAPSVODSGSLHNSSGSESTYCPONAGD 1081
QY 1081 LPSPDGVDYDODDYEDGALTSSGSVTFSSNSYGSOMSPDRGCVGTNNSGAYRFSSEGA 1140
Db 1082 LPSPDGVDYDODDYEDGALTSSGSVTFSSNSYGSOMSPDRGCVGTNNSGAYRFSSEGA 1141
QY 1141 QSSFEDESEEDFDSRFTDDELSTYRDSVYSCVTLPEYHSLYKKGGLMWMKRWCVLXD 1200
Db 1142 QSSFEDESEEDFDSRFTDDELSTYRDSVYSCVTLPEYHSLYKKGGLMWMKRWCVLXD 1201
QY 1201 ETLWFRSKOBAKQGMHLKKGGSSTLSRRNKKGMFVLROSKLMYFENDSEKUKGTV 1260
Db 1202 ETLWFRSKOBAKQGMHLKKGGSSTLSRRNKKGMFVLROSKLMYFENDSEKUKGTV 1261
QY 1261 EVRTAKEIIDNTKENGIDIIIMADRTPHLLAESPEASQWFSVLQVHASTDDEIQEMHD 1320
Db 1262 EVRTAKEIIDNTKENGIDIIIMADRTPHLLAESPEASQWFSVLQVHASTDDEIQEMHD 1321
QY 1321 EQANPQNAVGLDVLGIDSVCASDSPDRPNSFVITANRVLHGNADTPEEMHMITLLOR 1380
Db 1322 EQANPQNAVGLDVLGIDSVCASDSPDRPNSFVITANRVLHGNADTPEEMHMITLLOR 1381
QY 1381 SKGDTREVEGEFTVRGWLHKEVKNXSPMSSLKUKGMFVLTHNSLDVYKSSSEKNAKLG7 1440
Db 1382 SKGDTREVEGEFTVRGWLHKEVKNXSPMSSLKUKGMFVLTHNSLDVYKSSSEKNAKLG7 1441
QY 1441 LVNLSICSVVPPEBEKFKETGVNWNVYGRKHCVLRYTLKLNATRMSSVIOVVTOTKAP 1500
Db 1442 LVNLSICSVVPPEBEKFKETGVNWNVYGRKHCVLRYTLKLNATRMSSVIOVVTOTKAP 1501
QY 1501 IDPTFOOLIODIKENCINSNVEQIYKRNPLRVTYHHPHLSPLPLPYGGINNLKDKG 1560
Db 1502 IDPTFOOLIODIKENCINSNVEQIYKRNPLRVTYHHPHLSPLPLPYGGINNLKDKG 1561
QY 1561 YTTLODEBAIKFNLSLOQLESMSDPIPIQILOTGHDRLRLDELYCOLIKQTNKVPH9G 1620
Db 1562 YTTLODEBAIKFNLSLOQLESMSDPIPIQILOTGHDRLRLDELYCOLIKQTNKVPH9G 1621
QY 1621 SVGNLXSMOILLTCSGFLPSRGILKYLKFKHLKRIRBQFGTMEKXALFTYSLKKTTC 1680
Db 1622 SVGNLXSMOILLTCSGFLPSRGILKYLKFKHLKRIRBQFGTMEKXALFTYSLKKTTC 1681
QY 1681 REFVPSRDEIEALIHROEMSTVYCHGGSCKTINSHTTAGVEVEKLIRGLAMEDSRNM 1740
Db 1682 REFVPSRDEIEALIHROEMSTVYCHGGSCKTINSHTTAGVEVEKLIRGLAMEDSRNM 1741
QY 1741 FALFEYNGHVDKAIESTRTVADVLAKFEKLAATSEVGDLPWKFPKLYCFLDITDNVPKDS 1800
Db 1742 FALFEYNGHVDKAIESTRTVADVLAKFEKLAATSEVGDLPWKFPKLYCFLDITDNVPKDS 1801
QY 1801 VEEAFMEQAEHVIHGHHPAPEBNIOVLAALLOYLQOGYTLAAIPLLEEYYSIQRLK 1860
Db 1802 VEEAFMEQAEHVIHGHHPAPEBNIOVLAALLOYLQOGYTLAAIPLLEEYYSIQRLK 1861
QY 1861 ARIISQSTKTTPCEPERLEKRTSFLEGLTRSPFGTSVVRQKVEEOMLDMWIEEVSAR 1920
Db 1862 ARIISQSTKTTPCEPERLEKRTSFLEGLTRSPFGTSVVRQKVEEOMLDMWIEEVSAR 1921
QY 1921 ASIIDKRRKFOGNNQEOAMAKYALIKEMPGYSTLFDEYCKEGGEPQELMIGVSADVS 1980
Db 1922 ASIIDKRRKFOGNNQEOAMAKYALIKEMPGYSTLFDEYCKEGGEPQELMIGVSADVS 1981
QY 1981 VYKRGGRPLLEVQYEHILISFGAPLANTYKIVDEBELLETSEVVDVAKLMAIYSMTV 2040
Db 1982 VYKRGGRPLLEVQYEHILISFGAPLANTYKIVDEBELLETSEVVDVAKLMAIYSMTV 2041

QY 2041 KRRYSTTRSSASSQSSR 2057
Db 2042 KRRYSTTRSSASSQSSR 2058

RESULT 3
AAE11891
ID AAE11891 standard; Protein; 2048 AA.

AAE11891;

18-DEC-2001 (first entry)

Angiogenesis associated human myosin X2 (hMX2) protein variant.

Angiogenesis associated protein; AAP; cytosolic; cardiac; gene therapy;
ophthalmological; vulnery; myocardial infarction; macular degeneration;
diabetic retinopathy; angiogenesis; wound healing; prophylactic; vaccine;
rheumatoid arthritis; psoriasis; drug screening; tumour; transplantation;
cancer; therapeutic; diagnostic; human; myosin X2; MX2; mutant; mutein;
variant.

Homo sapiens.
Synthetic.

MO200170808-A2.

27-SEP-2001.

22-MAR-2001; 2001WO-US09609.

22-MAR-2000; 2000US-191134P.

(CURA-) CURAGEN CORP.
(GENT) GENENTECH INC.

Rastelli LK, Gerltzen M;

WPI: 2001-602775/68.

DR N-PSDB; AAD19120.

Novel angiogenesis associated polypeptides and polynucleotides encoding
the polypeptides, useful for modulating angiogenesis and for treating
tumors and cancers -

Claim 1; Page 42-49; 159pp; English.

The invention relates to angiogenesis associated proteins (AAP) and their
corresponding cDNA molecules, which are useful for modulating
angiogenesis. AAP proteins and nucleic acids are useful for promoting
wound healing, for example after organ transplantation, and in the
treatment of tumours, myocardial infarction, cancers, diabetic
retinopathy, macular degeneration, psoriasis and rheumatoid arthritis.
AAP proteins and DNA's are useful in potential prophylactic and
therapeutic applications implicated in a variety of disorders including
those related to angiogenesis, and also in diagnostic applications.
AAP cDNA is also useful in gene therapy. The invention also relates to
a method for screening a tissue sample for tumorigenic potential. AAP
proteins are used to screen drugs or compounds that modulate AAP activity
or expression as well as treating disorders characterized by insufficient
or excessive production of AAP or production of AAP forms that have
decreased or aberrant activity compared to the wild type protein, or
modulate biological function that involve AAP. The present sequence
is human myosin X2 (hMX2) variant protein which is an angiogenesis
associated protein (AAP) of the invention. hMX2 is upregulated in the in
vitro model of angiogenesis and is likely to be involved in transducing
signals from angiogenic factors, perhaps modulating the cytoskeleton.

Sequence 2048 AA;

Query Match 95.5%; Score 10247; DB 22; Length 2048;
Best Local Similarity 99.8%; Pred. No. 0;

Matches 1961; Conservative 1; Mismatches 2; Indels 0; Gaps 0;			
Qy	1	FCLOGRVVLRENGQHFPSTVNSCAGIYVFRDYGQVFFYYKOSTITHQKTMHPETNEE	60
Db	1	FCLOGRVVLRENGQHFPSTVNSCAGIYVFRDYGQVFFYYKOSTITHQKTMHPETNEE	60
Qy	61	GVDDMASTLTELHGGSIMYNLFQRYKKNQIWTYIGSILASVNPQIPIAGIYEPATMEQYSR	120
Db	61	GVDDMASTLTELHGGSIMYNLFQRYKKNQIWTYIGSILASVNPQIPIAGIYEPATMEQYSR	120
Qy	121	RHLGELPPIHFAIANECYRCMLKRNOCCLIKGSGAGCTESTKILKFLSVISQOQSLB	180
Db	121	RHLGELPPIHFAIANECYRCMLKRNOCCLIKGSGAGCTESTKILKFLSVISQOQSLB	180
Qy	181	LSLKEKTSQVERALILESSPIMEAFNAKTYVNNNSRFGIFVQJNICQKNGIIGGRIYVC	240
Db	181	LSLKEKTSQVERALILESSPIMEAFNAKTYVNNNSRFGIFVQJNICQKNGIIGGRIYVC	240
Qy	241	ILSSQNRVVRQNGERNYHIFVALLAGLEHEEREFEYLSIPENYHYLNOGCVEDKTIISD	300
Db	241	ILSSQNRVVRQNGERNYHIFVALLAGLEHEEREFEYLSIPENYHYLNOGCVEDKTIISD	300
Qy	301	QSSFRVITAMDMVQMSKEVREVSRLAGILHLSNIEPITTAGAQSFPKTALGREAELL	360
Db	301	QSSFRVITAMDMVQMSKEVREVSRLAGILHLSNIEPITTAGAQSFPKTALGREAELL	360
Qy	361	GLDPTQLTALTORSMPLRGEELITPLNQOAVDSRDSIALMALYACCFEVIKIKISRIK	420
Db	361	GLDPTQLTALTORSMPLRGEELITPLNQOAVDSRDSIALMALYACCFEVIKIKISRIK	420
Qy	421	GNEDFKSGILDI FGEPENFVNHFEOFNINYANEKLOEYFNKHI FSLQOLEYSGREGIYWE	480
Db	421	GNEDFKSGILDI FGEPENFVNHFEOFNINYANEKLOEYFNKHI FSLQOLEYSGREGIYWE	480
Qy	481	DIIDMINGECDLIEKKGLLALINEESHFPQATDSTLLEKLSQANHHFYKPRVAVN	540
Db	481	DIIDMINGECDLIEKKGLLALINEESHFPQATDSTLLEKLSQANHHFYKPRVAVN	540
Qy	541	NGGVKIRAGEVOYDVAGILEKONDTFRDPLNLRSRPDIYDLFEHVSRRNOOTLKC	600
Db	541	NGGVKIRAGEVOYDVAGILEKONDTFRDPLNLRSRPDIYDLFEHVSRRNOOTLKC	600
Qy	601	GSKHRRPTVVSQFQVDSLHSLMATLSSNPFVRCIKPMQKMPDQDAVVLNQLRYSG	660
Db	601	GSKHRRPTVVSQFQVDSLHSLMATLSSNPFVRCIKPMQKMPDQDAVVLNQLRYSG	660
Qy	661	MEETVIRIRAGAVRRPFODFYGRYKVLNMLNLPEDVFGKCTSLQLYDASNSEWQLGK	720
Db	661	MEETVIRIRAGAVRRPFODFYGRYKVLNMLNLPEDVFGKCTSLQLYDASNSEWQLGK	720
Qy	721	TKVFLRESLLEOKLEKREEVSHAAVYIRAHVYGLARQYRVLYCWTIOKNYAPFL	780
Db	721	TKVFLRESLLEOKLEKREEVSHAAVYIRAHVYGLARQYRVLYCWTIOKNYAPFL	780
Qy	781	RRRFLHLKKAAYVFOQLRGQIARVYROLLEKREOEKKQOESEKKKRESEERER	840
Db	781	RRRFLHLKKAAYVFOQLRGQIARVYROLLEKREOEKKQOESEKKKRESEERER	840
Qy	841	ERRREALRQOEETKQOELALOKSOKEALJTRLEHCQENKQYBELTRLEKEIEDIQ	900
Db	841	ERRREALRQOEETKQOELALOKSOKEALJTRLEHCQENKQYBELTRLEKEIEDIQ	900
Qy	901	RMKEQOELSTLTSLOKLOERRDOELRLLEBA CRAAOHFLBSLNDEIDECVNRTERSL	960
Db	901	RMKEQOELSTLTSLOKLOERRDOELRLLEBA CRAAOHFLBSLNDEIDECVNRTERSL	960
Qy	961	SGGSEFSSEIASEACEKPNFNSQPYPEEVEDEGEAUNDADAKSPNPSHSHSDQRTS	1020
Db	961	SGGSEFSSEIASEACEKPNFNSQPYPEEVEDEGEAUNDADAKSPNPSHSHSDQRTS	1020
Qy	1021	GRTSDSDSEEDPYNDVTVPISPSADSTVLLAPSVQDSGLSNSSGSSSTYMPONAGD	1080
Db	1021	GRTSDSDSEEDPYNDVTVPISPSADSTVLLAPSVQDSGLSNSSGSSSTYMPONAGD	1080

Qy	1081	LPSPDGDYDYDDYEDGALITSGSSVTSNSYSGQMSPDYRCSVGTINSSGAYRESSBGA	1140
Db	1081	LPSPDGDYDYDDYEDGALITSGSSVTSNSYSGQMSPDYRCSVGTINSSGAYRESSBGA	1140
Qy	1141	QSSFEDSEEDPSRPTDDELSTYRDSYSCVTLPLFYHSLYMKGLMNSWRKRCVCLKD	1200
Db	1141	QSSFEDSEEDPSRPTDDELSTYRDSYSCVTLPLFYHSLYMKGLMNSWRKRCVCLKD	1200
Qy	1201	ETFLMFRSQOZALKQGMHLKGGSSSTLSRRWKKRWFLRQSKLMPYENDSEELKGVN	1260
Db	1201	ETFLMFRSQOZALKQGMHLKGGSSSTLSRRWKKRWFLRQSKLMPYENDSEELKGVN	1260
Qy	1261	EVRTAKEIINDNTKNGIDIIIMADRTPHLIAESPDAQSFSLQVHASTQOEIOEMND	1320
Db	1261	EVRTAKEIINDNTKNGIDIIIMADRTPHLIAESPDAQSFSLQVHASTQOEIOEMND	1320
Qy	1321	EQANPQNAVGLIDVGLIDSVCASDSPDRPNSFVITIANRVLHCNADTPREMHWTLLQR	1380
Db	1321	EQANPQNAVGLIDVGLIDSVCASDSPDRPNSFVITIANRVLHCNADTPREMHWTLLQR	1380
Qy	1381	SKGDRVREGOEIRVGMHLKEVKNSPKMSLKLKRWFLVLTNLSLDYKSSRKMLXGCT	1440
Db	1381	SKGDRVREGOEIRVGMHLKEVKNSPKMSLKLKRWFLVLTNLSLDYKSSRKMLXGCT	1440
Qy	1441	LVNLSLCSVPPDEKIFKETGYMNTVYGRKHCYRLYTQLINEATRMSVIONVDTKAP	1500
Db	1441	LVNLSLCSVPPDEKIFKETGYMNTVYGRKHCYRLYTQLINEATRMSVIONVDTKAP	1500
Qy	1501	IDTPTQQLIODIKENCLNSDVEQIKRNPILRYTHHPHLSPLPLPYGDIMNLNLLKDG	1560
Db	1501	IDTPTQQLIODIKENCLNSDVEQIKRNPILRYTHHPHLSPLPLPYGDIMNLNLLKDG	1560
Qy	1561	YTTLODEAKIKINSIQOLESMSDPIPIIGLIQTHDLRPLDELYCOLIKQTNKVPHG	1620
Db	1561	YTTLODEAKIKINSIQOLESMSDPIPIIGLIQTHDLRPLDELYCOLIKQTNKVPHG	1620
Qy	1621	SVGNLYSMOILCTCLCTFLPSRGIKLYLKFHLKRIREOPGTMEKVALFTYESLKTTC	1680
Db	1621	SVGNLYSMOILCTCLCTFLPSRGIKLYLKFHLKRIREOPGTMEKVALFTYESLKTTC	1680
Qy	1681	REFVPSRDEIALIHRQEWSTVYCHGGSCKITINSHTAGVYEKILRGLAMEDSRNM	1740
Db	1681	REFVPSRDEIALIHRQEWSTVYCHGGSCKITINSHTAGVYEKILRGLAMEDSRNM	1740
Qy	1741	FALFENGVHVDCAISRTYVADVLAKEFLAATSSEVDLPKMFYKLYFLDLDNVPKOS	1800
Db	1741	FALFENGVHVDCAISRTYVADVLAKEFLAATSSEVDLPKMFYKLYFLDLDNVPKOS	1800
Qy	1801	VEFAFMFEQAHBAVYHGHHPAPEENLOVLAALRLQYLODYYTLHAIPLEEVYSLQRLK	1860
Db	1801	VEFAFMFEQAHBAVYHGHHPAPEENLOVLAALRLQYLODYYTLHAIPLEEVYSLQRLK	1860
Qy	1861	ARISQSTKTFPCERLEKRRISFLEGTLLRSFRTSVVRQYVEECMLDMMIKEEVSAR	1920
Db	1861	ARISQSTKTFPCERLEKRRISFLEGTLLRSFRTSVVRQYVEECMLDMMIKEEVSAR	1920
Qy	1921	ASIIDKMRKFGQMNQOAMAKYMALIKEMPGGSLTFVYCEGEG	1984
Db	1921	ASIIDKMRKFGQMNQOAMAKYMALIKEMPGGSLTFVYCEGEG	1984

RESULT 4
 ID ABG10631 standard; Protein; 2408 AA.
 AC ABG10631;
 DT 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #10622.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM	food supplement; medical imaging; diagnostic; genetic disorder.
XX	
OS	Homo sapiens.
XX	
PN	WO200175067-A2.
XX	
PD	11-OCT-2001.
XX	
PF	30-MAR-2001; 2001WO-US08631.
XX	
PR	31-MAR-2000; 2000US-0540217.
XX	
PR	23-AUG-2000; 2000US-0649167.
XX	
PA	(HYSE-) HYSEQ INC.
XX	
PI	Dymanac RT, Liu C, Tang YT;
XX	
PI	WPI, 2001-639362/73.
DR	N-PSDB; AAS74818.
XX	
PT	New isolated polypeptide and encoded polypeptides, useful in
XX	
PT	diagnostics, forensics, gene mapping, identification of mutations
XX	
PT	responsible for genetic disorders or other traits and to assess
XX	
PT	biodiversity
XX	
PS	Claim 20; SEQ ID No 40990; 103pp; English.
XX	
CC	The invention relates to isolated polynucleotide (I) and
XX	
CC	polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX	
CC	polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX	
CC	and gene mapping, and in recombinant production of (II). The
XX	
CC	polynucleotides are also used in diagnostics as expressed sequence tags
XX	
CC	for identifying expressed genes. (I) is useful in gene therapy techniques
XX	
CC	to restore normal activity of (II) or to treat disease states involving
XX	
CC	(II). (II) is useful for generating antibodies against it, detecting or
XX	
CC	quantitating a polypeptide in tissue, as molecular weight markers and as
XX	
CC	a food supplement. (II) and its binding partners are useful in medical
XX	
CC	imaging of sites expressing (II). (I) and (II) are useful for treating
XX	
CC	disorders involving aberrant protein expression or biological activity.
XX	
CC	The polypeptide and polynucleotide sequences have applications in
XX	
CC	diagnostics, forensics, gene mapping, identification of mutations
XX	
CC	responsible for genetic disorders or other traits to assess biodiversity
XX	
CC	and to produce other types of data and products dependent on DNA and
XX	
CC	amino acid sequences. ABG00010-ABG30377 represent novel human
XX	
CC	diagnostic amino acid sequences of the invention.
XX	
CC	Note: The sequence data for this patent did not appear in the printed
XX	
CC	specification, but was obtained in electronic format directly from WIPO
XX	
XX	at ftp.wipo.int/pub/published_pct_sequences.
XX	
SO	Sequence 2408 AA;
XX	
QY	Query Match 66.8%; Score 716615; DB 22; Length 2408;
XX	
QY	Best Local Similarity 90.3%; Pred. No. 0;
XX	
QY	Matches 1408; Conservative 23; Mismatches 52; Indels 77; Gaps 10;
XX	
QY	555 VRGILEKRDVFRDDLNL-----LIRSRFFIYDLFEHVSSRNNDTLKCG----- 601
XX	
QY	214 I PGLIQE----FRDALNMGPDKEIFKETELVNDMDKINGRIERAASATCEGLRIIV 269
XX	
QY	602 -----SGRRRTYSQ-----FKVDSLHSLMATLSSSNPFVRCIK--- 637
XX	
QY	270 VSPFHSVHLFTNVGPQVRNRIACACERAGMEYRSASLSLNAVAEG---YVREKEIRA 326
XX	
QY	638 -----PNM-----QKMPDQDOAVVLNQLYSGSLQETVRIKKGYAVRRRFQDFY 682
XX	
QY	327 AEKTTTNGNPDPDPDCCRRHPDQFDQAVVLNQLYSGSLQETVRIKKGY----- 375
XX	
QY	683 KRYKVLNMLALPEDVRGKCTSLDLYASNSNEMOLGKTKVFLRSLQKLEKREBEVS 742
XX	
QY	376 --YKVLNMLNLPEDVRGKCTSLDLYASNSNEMOLGKTKVFLRSLQKLEKREBEVS 433
XX	
QY	743 HAANVIRAHVIGFLARKQYRKLYCVVIIQKXYRAPELLRRFLHKKAAIYFQKLRQI 802

Dh	433	HAANVIRAHVGLFARQYKRVLCVILIQKNVAFILRRRFLHLKKAALVYFQQLAGQI	499
Oy	803	ARRVYROLLAEKRBOEKKQOEBEKKKKRBEEREREREREAFLRAQOEETRKOOELE	862
Dh	494	ARRVYROLLAEKRQOEKKQOEBEKKKKRBEEREREREREAFLRAQOEETRKOOELE	553
Oy	863	ALQSOKEAEITRELEKOKENKQOEELTRLEKEIEDIQRNKEOQELTLEASLOKLEOR	922
Dh	554	ALQSOKEAEITRELEKOKENKQOEELTRLEKEIEDIQRNKEOQELTLEASLOKLEOR	613
Oy	923	DOELRLREBEACRAQOEFLESINDEIDECVRNERSLSGSESESSLEIABESACEKNFN	982
Dh	614	DOELRLREBEACRAQOEFLESINDEIDECVRNERSLSGSESESSLEIABESACEKNFN	673
Oy	983	FSQOY P --- BEEYDEGEADDDAFKOSPNESEHGSDQRTSGIRTSDDSEEDPYMNDT	1038
Dh	674	LSKRPNPKRKOEDVEGEADDDAFKOSPNESEHGSDQRTSGIRTSDDSEEDPYMNDT	733
Oy	1039	VVPFSPADSTVLLABVOPDGSILHNSSEESTYCMQONNGDLPSPGDYDODDQYDEG	1098
Dh	734	VVPFSPADSTVLLABVOPDGSILHNSSEESTYCMQONNGDLPSPGDYDODDQYDEG	793
Oy	1099	AITGSSSVTFNSNGSQMSFPDYRCVSTYNSGAYRFSSEGAQSPEDSEEDPFSRPTD	1158
Dh	794	AITGSSSVTFNSNGSQMSFPDYRCVSTYNSGAYRFSSEGAQSPEDSEEDPFSRPTD	853
Oy	1159	DELSYRDSVYSCVTLPEFHSFLYMKGGLNMSWRKRCVCLKDETFLFRSKOELKQGM	1218
Dh	854	DELSYRDSVYSCVTLPEFHSFLYMKGGLNMSWRKRCVCLKDETFLFRSKOELKQGM	913
Oy	1219	HKKGSGSTLSRRRWKRWVFLRQSKLMPENDBEELKGVETRTAKEIIDNTTKNGI	1278
Dh	914	HKKGSGSTLSRRRWKRWVFLRQSKLMPENDBEELKGVETRTAKEIIDNTTKNGI	973
Oy	1279	DIINADRTPHLIAESPEDASOMFVLSQVASTDOELQEMHDEQANQONAVGTLDVGLID	1338
Dh	974	DIINADRTPHLIAESPEDASOMFVLSQVASTDOELQEMHDEQANQONAVGTLDVGLID	1033
Oy	1339	SVCSASDPDRNSFVLIITANRVLHCNADTPEBMHMITLLQORSKGDTRVEGOEIVRGM	1398
Dh	1034	SVCSASDPDRNSFVLIITANRVLHCNADTPEBMHMITLLQORSKGDTRVEGOEIVRGM	1093
Oy	1399	HKEVNSPKWSSSLKKRWVFLTNSLDYKSSSKN-ALKLGTIVLNSLCSVPPDEKIF	1457
Dh	1094	HKEVNSPKWSSSLKKRWVFLTNSHYKSSSEKONRSNMGTIVLNSLCSVPPDEKIF	1153
Oy	1458	KETGYMWVYVGRGCRILYTKLINBATRNSSVQNTDTPKAPIDTPTOOLIOIKENCL	1517
Dh	1154	KETGYMWVYVGRGCRILYTKLINBATRNSSAIQNTDTPKAPIDTPTOOLIOIKENCL	1213
Oy	1518	NSDVAVEOIKYKRNPIRLRTHNPLHSPLPLRPGYDINLMLDKGYTTLQDEAKIFNSLQ	1577
Dh	1214	NSDVAVEOIKYKRNPIRLRTHNPLHSPLPLRPGYDINLMLDKGYTTLQDEAKIFNSLQ	1273
Oy	1578	LESMSDPIPIIOGILQIGHDLRPLRDELXCOLIKQTNKVPHPGSGVULYSQOILTCISCT	1637
Dh	1274	LESMSDPIPIIOGILQIGHDLRPLRDELXCOLIKQTNKVPHPGSGVULYSQOILTCISCT	1333
Oy	1638	FLPBRGLIKYLKFLKRIROFPETEMEKALFVYESLXKTKCPEFSPSDEIEALIHRO	1697
Dh	1334	FLPBRGLIKYLKFLKRIROFPETEMEKALFVYESLXKTKCPEFSPSDEIEALIHRO	1393
Oy	1698	EMTSTYVCHGGGSKITINTSHTTAGEVEKILRGLAEDSRNMALFEYNGHVDAIESR	1757
Dh	1394	EMTSTYVCHGGGSKITINTSHTTAGEVEKILRGLAEDSRNMALFEYNGHVDAIESR	1453
Oy	1758	TVVADVLAKTEKLAATSEVDDLPMKFYFKLYCFLDTONVPCDSYEFAPFMEQAEAVIHG	1817
Dh	1454	TVVADVLAKTEKLAATSEVDDLPMKFYFKLYCFLDTONVPCDSYEFAPFMEQAEAVIHG	1513
Oy	1818	HHPRPEENLOYLAALRIQYIQOGDYTLHLAALPRLBEVYSLOLRKARISQSTTFPCERLE	1877
Dh	1514	HHPRPEENLOYLAALRIQYIQOMDYTLQAAALPRLBEVYSLOLRKARISQSTTFPCERLE	1573

QY 1878 KRTSFLBGLRLRSFRTGSVVRQVEEQMLDMWIMEEVSASASIIDMKRFGQNNQO 1937
DB 1574 KRTSFLBGLRLRSFRTGSVVRQVEEQMLDMWIMEEVSASASIIDMKRFGQNNQO 1633
QY 1938 AAKAYVALIKEMPGYSTLFDVECKGKGGFPOELMLGVADANSVYRGRGRPLEVFQYH 1997
DB 1634 AAKAYVALIKEMPGYSTLFDVECKGKGGFPOELMLGVADANSVYRGRGRPLEVFQYH 1693
QY 1998 ILSFGAPLANTYIVVDERELFETSEBVVDVAKLMKAYISMIVKXRYSTTRASAGGSSR 2057
DB 1694 ILSFGAPLANTYIVVDERELFETSEBVVDVAKLMKAYISMIVKXRYSTTRASAGGSSR 1753

RESULT 5
ABG10624
ID ABG10624 standard; Protein; 1515 AA.
XX
AC ABG10624;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #10615.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US06631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR N-PSDB; AAS74811.
XX
PT MPI; 2001-639362/73.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 40983; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pat_sequences.

XX
SQ Sequence 1515 AA;
Query Match 63.1%; Score 6775.5; DB 22; Length 1515;
Best Local Similarity 93.3%; Pred. No. 0;
Matches 1342; Conservative 6; Mismatches 8; Indels 81; Gaps 5;
QY 641 QMPPQDPAVVLNOLARYSGMLETIRIRKAGYAVRRPQDPYKRYKYLNRNALPEDIYRG 700
DB 180 RMPPQDPAVVLNOLARYSGMLETIRIRKAGYAVRRPQDPYKRYKYLNRNALPEDIYRG 226
QY 701 KTSLLQLYDASNSWQGLKTVPLRESLEQLERREBEVSHAMVIRAHVGLFLARKQ 760
DB 227 KTSLLQLYDASNSWQGLKTVPLRESLEQLERREBEVSHAMVIRAHVGLFLARKQ 286
QY 761 YKRVLYCVVITQXNRAFLRRRFLTKKAAIVPQKQGLARVYQQLAEKREOEK 820
DB 287 YKRVLYCVVITQXNRAFLRRRFLTKKAAIVPQKQGLARVYQQLAEKREOEK 346
QY 821 KQEEEEKKKREEREREREREREAELAQQEETRKOQELALQSQKEAELTRELKQ 880
DB 347 KQEEEEKKKREEREREREREREAELAQQEETRKOQELALQSQKEAELTRELKQ 406
QY 881 KENKQVEETLRLEKEIEDLQRMKEQOELSTFASLQKLQERRDQELRLLEBACPAQEF 940
DB 407 KENKQVEETLRLEKEIEDLQRMKEQOELSTFASLQKLQERRDQELRLLEBACPAQEF 466
QY 941 LESLNFDEIDECVNRTERSLSGGSFSSSELAESACEKPNFNSQPYPEEYDEGFELAD 1000
DB 467 LESLNFDEIDECVNRTERSLSGGSFSSSELAESACEKPNFNSQPYPEEYDEGFELAD 526
QY 1001 DAFKDSPNPSEHGHDDQRTSGIRTSDDSEEDPYNNDTVPVPSPADSTVLLAPVQDSG 1060
DB 527 DAFKDSPNPSEHGHDDQRTSGIRTSDDSEEDPYNNDTVPVPSPADSTVLLAPVQDSG 568
QY 1061 SLHNSSGESTYCMPQNAAGDLPSPDGDYDDYDDYEDGATISGSSVTSNSYGSQMSPDY 1120
DB 569 SLHNSSGESTYCMPQNAAGDLPSPDGDYDDYDDYEDGATISGSSVTSNSYGSQMSPDY 606
QY 1121 RCVSTYNSSGAYRFSSEGAQSSFEDESDPSPRPTDDELRYRDSVYSCVTLVYFHSF 1180
DB 607 RCVSTYNSSGAYRFSSEGAQSSFEDESDPSPRPTDDELRYRDSVYSCVTLVYFHSF 666
QY 1181 LYMKGLMNSWGRKRCVLTDEFTLFRSKQELKKGWMLHKKGSGSTLSRRMKKRWFL 1240
DB 667 LYMKGLMNSWGRKRCVLTDEFTLFRSKQELKKGWMLHKKGSGSTLSRRMKKRWFL 726
QY 1241 ROSKLMYFENDSEKLGTVFVRTAKEIIDNTKENGIDIIWADRTPHLIABSPEDASQW 1300
DB 727 ROSKLMYFENDSEKLGTVFVRTAKEIIDNTKENGIDIIWADRTPHLIABSPEDASQW 786
QY 1301 FSVLSQVHAASSTQELQEHMDEQANPQAVGTLIDVLISVCASDSFDRPNSVYIITANRV 1360
DB 787 FSVLSQVHAASSTQELQEHMDEQANPQAVGTLIDVLISVCASDSFDRPNSVYIITANRV 846
QY 1361 LHCNADPPEMHWITTLQORSKGDTRVGEQETVGMVHKEVKNSEPKMSLTLKRWFL 1420
DB 847 LHCNADPPEMHWITTLQORSKGDTRVGEQETVGMVHKEVKNSEPKMSLTLKRWFL 906
QY 1421 THNSLDYKSSBKNAIKLGLVLNLSGVPPDEKIFKETGYMNTVYGRKCYRLYTQL 1480
DB 907 THNSLDYKSSBKNAIKLGLVLNLSGVPPDEKIFKETGYMNTVYGRKCYRLYTQL 966
QY 1481 LNEARWSSVIONVDTKAPITPTPQOILQDIKENCASDVVEQIYKRNPIRLRTHPLH 1540
DB 967 LNEARWSSVIONVDTKAPITPTPQOILQDIKENCASDVVEQIYKRNPIRLRTHPLH 1026
QY 1541 SPLPLPFGDILNLTNKKGYTTLQDEAIKFNLSQQLESMSDPIPIIGIQLQTHDLP 1600
DB 1027 SPLPLPFGDILNLTNKKGYTTLQDEAIKFNLSQQLESMSDPIPIIGIQLQTHDLP 1086
QY 1601 LADBLVQCLIKQTNKVPHPGSGVNLVSWQILTGLSCTPLPSRGIKLYLPHLKRIREQFP 1660
DB 1086 LADBLVQCLIKQTNKVPHPGSGVNLVSWQILTGLSCTPLPSRGIKLYLPHLKRIREQFP 1660

Db 1087 LRDELVYCOLIKQNNKVPHPGSGVNLVSMQILTCLSCFTPLPSRGILKYLKFKHLKRIEOPF 1146
Qy 1661 GTEMEKXALFTYESLKKTKCREFPVPSRDELEALIHREMTSTVYCHGGGCKITINSHTT 1720
Db 1147 GTEMEKXALFTYESLKKTKCREFPVPSRDELEALIHREMTSTVYCHGGGCKITINSHTT 1206
Qy 1721 AGEVEKILIRGLAMEDSRNMFALFEYNGHDKAIESRTVADVLAKEKLAATSEVGDLP 1780
Db 1207 AGEVEKILIRGLAMEDSRNMFALFEYNGHDKAIESRTVADVLAKEEN-ALTKL- 1261
Qy 1781 WKFEYKLYCFLDNDNVKDSVEFAFMEOHNAVHGHHPAPRENQVLAALQYLOQD 1840
Db 1262 -----LTVSF-----QAEHAVIHGHHPAPRENQVLAALQYLOQD 1298
Qy 1841 YTLHAAIPLEEVYSIORLKARISQSTKTFTPCERLEKRTSFLEGLTRSPRTGSVMQ 1900
Db 1299 YTLHAAIPLEEVYSIORLKARISQSTKTFTPCERLEKRTSFLEGLTRSPRTGSVMQ 1358
Qy 1901 KVEEOMLDMWIKKEVSSARASIIDKWRKFGQNNQOQAKYMALIKEMPGYSTLPDVE 1960
Db 1359 KVEEOMLDMWIKKEVSSARASIIDKWRKFGQNNQOQAKYMALIKEMPGYSTLPDVE 1418
Qy 1961 CKRGFPQELMLGVSDAVSVYKRGEPLEVFQYEHILSFQAPLANTYKIIVDERELLF 2020
Db 1419 CKRGFPQELMLGVSDAVSVYKRGEPLEVFQYEHILSFQAPLANTYKIIVDERELLF 1478
Qy 2021 ETESEVVDVAKLMKAYISMIVKKRYSTTRSSASQSSSR 2057
Db 1479 ETESEVVDVAKLMKAYISMIVKKRYSTTRSSASQSSSR 1515

RESULT 6
AAU17082
ID AAU17082 standard; Protein: 1031 AA.
XX AAU17082;
DT 07-NOV-2001 (first entry)
XX
DE Novel signal transduction pathway protein, Seq ID 647.
XX
KW Neuroprotective; cytoskeletal; dermatological; immunosuppressive; tumour;
KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anemia; hyperproliferative disorder; Gaucher's disease;
KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS;
KW acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN MO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-US01312.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 26-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.

Db 786 EAVIHGHHPAPEENLQVLAALRLQYLOGDYTLHAIPPLEBYVSLQRLKARISQSTKTFT 845
QY 1872 PCRLEKRTSPLEGLTRRSFRTGSVVRQVVEEOMLDMWIKKEVSSARASIIDKMRKFO 1931
Db 846 PCRLEKRTSPLEGLTRRSFRTGSVVRQVVEEOMLDMWIKKEVSSARASIIDKMRKFO 905
QY 1932 GMMQEQAMAKYMLIKEMPGYSTLPDVECKEGFPQELMWGVSADAVSYVKRGEGRPLE 1991
Db 906 GMMQEQAMAKYMLIKEMPGYSTLPDVECKEGFPQELMWGVSADAVSYVKRGEGRPLE 965
QY 1992 VFOYEHLSFGAPLANTYKIVDERELFTSEVVDVAKLMKAYISMIVKKRYSTTRSSAS 2051
Db 966 VFOYEHLSFGAPLANTYKIVDERELFTSEVVDVAKLMKAYISMIVKKRYSTTRSSAS 1025
QY 2052 SQGSSR 2057
Db 1026 SQGSSR 1031

RESULT 7
AAM93762 standard; Protein; 984 AA.
ID AAM93762
XX AAM93762;
AC AAM93762;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 3757.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JUL-0194486.
XX
PR 11-JUN-2000; 2000JUL-0118774.
XX
PR 02-MAY-2000; 2000MAY-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Oca T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
XX
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
DR N-PSDB; AAK94716.
XX
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
XX
PT use in genetic manipulation -
XX
PS
XX
PS Claim 8; SEQ ID NO 3757; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
XX
CC clones. 830 cDNA molecules encoding a human protein have been
XX
CC isolated and nucleotide sequences of 5' and 3'-ends of the cDNA
XX
CC molecules have been determined. Primers for synthesizing the full length
XX
CC cDNA are useful for clarifying the function of the protein encoded by
XX
CC the cDNA. The full length clones were obtained by construction of full
XX
CC length enriched cDNA libraries that were synthesised by the oligo-capping
XX
CC method. The primers enable the production of the full length cDNA easily
XX
CC without any special methods. The present sequence is a polypeptide
XX
CC encoded by a full length human cDNA of the invention.
XX
CC Note: The sequence data for this patent did not form part of the printed
XX
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
XX
XX Sequence 984 AA:

Query Match 48.1%; Score 5166; DB 22; Length 984;
Best Local Similarity 99.6%; Pred. No. 0;

Matches 980; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
QY 1074 MQONAGDLPSPGDYDDDDYEDGATITSGSVTFPSNLSGSGMSDPYCSVGTYNSSGAY 1133
Db 1 MQONAGDLPSPGDYDDDDYEDGATITSGSVTFPSNLSGSGMSDPYCSVGTYNSSGAY 60
QY 1134 RFSSEGAOSSFEDSEEDFSDRFTDDELSTYRDSVYSCVTLPEFHSFLYMKGLMNSWR 1193
Db 61 RFSSEGAOSSFEDSEEDFSDRFTDDELSTYRDSVYSCVTLPEFHSFLYMKGLMNSWR 120
QY 1194 RMCVTLKDETFILFRSKQELKQGMHLKGGSSSTLSRRMKRWLVLSQSLMYEENDE 1253
Db 121 RMCVTLKDETFILFRSKQELKQGMHLKGGSSSTLSRRMKRWLVLSQSLMYEENDE 180
QY 1254 EKLKGTVEVRTAKEIINDTTKENGIDIMADSTPHLIAESPDAQMSVLSQVASTDQ 1313
Db 181 EKLKGTVEVRTAKEIINDTTKENGIDIMADSTPHLIAESPDAQMSVLSQVASTDQ 240
QY 1314 EIQEMHDEQANPQNAVGLDVLIDSVCSADSPDRPNSFVITANRVLHCNADTPEEMH 1373
Db 241 EIQEMHDEQANPQNAVGLDVLIDSVCSADSPDRPNSFVITANRVLHCNADTPEEMH 300
QY 1374 WITLLQRSKGDTRVGEQETVARGWLHKEYKNSLKLKRWVLTNLSLDYKSSSK 1433
Db 301 WITLLQRSKGDTRVGEQETVARGWLHKEYKNSLKLKRWVLTNLSLDYKSSSK 360
QY 1434 NALKLGLVLNLSLGSVPPDEKIFKETGYMNVTYGRKHCVRLYTKLNEATRMSAVION 1493
Db 361 NALKLGLVLNLSLGSVPPDEKIFKETGYMNVTYGRKHCVRLYTKLNEATRMSAVION 420
QY 1494 VTDTRAPIDTPTQOILQDIKENCNSDVVEQIKYKNPILRYTHPLHSLPLPYGDIWL 1553
Db 421 VTDTRAPIDTPTQOILQDIKENCNSDVVEQIKYKNPILRYTHPLHSLPLPYGDIWL 480
QY 1554 NLLKQKGYTLLODEAKIFNSIQLESMSDPIPIIQGLQTHDRLPRDELVCOLIKQT 1613
Db 481 NLLKQKGYTLLODEAKIFNSIQLESMSDPIPIIQGLQTHDRLPRDELVCOLIKQT 540
QY 1614 NKVPHGVSQNLYSQOILTCLSCTFLPBSGLIKYLPFLKRIREQPGTEMEKVALFYE 1673
Db 541 NKVPHGVSQNLYSQOILTCLSCTFLPBSGLIKYLPFLKRIREQPGTEMEKVALFYE 600
QY 1674 SLKTKRCREFPVSRDEIEALIHROEMTSTVYCHGGGSCKITINSHTTAGEVEVKLIRGLA 1733
Db 601 SLKTKRCREFPVSRDEIEALIHROEMTSTVYCHGGGSCKITINSHTTAGEVEVKLIRGLA 660
QY 1734 MEDSRNMFALFEYNQHVDAIESRTVADVLAKEFKLAATSEVGLPMKEFYKLYRFLDT 1793
Db 661 MEDSRNMFALFEYNQHVDAIESRTVADVLAKEFKLAATSEVGLPMKEFYKLYRFLDT 720
QY 1794 DNVPKDSVEFAFMFOAHEAVTGHHPAPEENLQVLAALRLQYLOGDYTLHAIPPLEBV 1853
Db 721 DNVPKDSVEFAFMFOAHEAVTGHHPAPEENLQVLAALRLQYLOGDYTLHAIPPLEBV 780
QY 1854 YSLQRLKARISQSTKTFTPCRLEKRTSPLEGLTRRSFRTGSVVRQVVEEOMLDMWIK 1913
Db 781 YSLQRLKARISQSTKTFTPCRLEKRTSPLEGLTRRSFRTGSVVRQVVEEOMLDMWIK 840
QY 1914 EVSSARASIIDKMRKFOGMMQEQAMAKYMLIKEMPGYSTLPDVECKEGFPQELMWG 1973
Db 841 EVSSARASIIDKMRKFOGMMQEQAMAKYMLIKEMPGYSTLPDVECKEGFPQELMWG 900
QY 1974 VSADAVSYVKRGEGRPLEVFOYEHLSFGAPLANTYKIVDERELFTSEVVDVAKLMK 2033
Db 901 VSADAVSYVKRGEGRPLEVFOYEHLSFGAPLANTYKIVDERELFTSEVVDVAKLMK 960
QY 2034 AYISMIVKKRYSTTRSSASQSSR 2057
Db 961 AYISMIVKKRYSTTRSSASQSSR 984

RESULT 8
ABG25601

Db	1390	ALHROEWSTFVCHGGGSKCITINSHTT	1418
RESULT 9			
AAAB42020			
ID	AAAB42020	standard; Protein; 869 AA.	
XX			
AC	AAAB42020;		
XX			
DT	08-FEB-2001	(first entry)	
XX			
DE	Human ORFX ORF1784 polypeptide sequence SEQ ID NO:3568.		
XX			
KM	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;		
KM	vulnerary; antipapillary; antiparkinsonian; nootropic; neuroprotective;		
KM	immunocytotoxic; osteopontin; antihypertensive; immunosuppressive; cardiant;		
KM	immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;		
KM	hypotensive; dermatological; immunosuppressive; anti-inflammatory;		
KM	antiviral; antibacterial; antifungal; antirheumatic; antichyroid;		
KM	antianemic; gene therapy; cancer; proliferative disorder; hypertension;		
KM	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
KM	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;		
KM	cholesterol ester storage; systemic lupus erythematosus; infection;		
KM	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
KM	allergy; aplastic anaemia; nocturnal hemoglobinuria; burn; wound;		
KM	bone damage; cartilage damage; antiinflammatory disease; coagulation;		
KM	thrombosis; contraceptive.		
XX			
OS	Homo sapiens.		
XX			
PN	WO2000058473-A2.		
PD			
XX	05-OCT-2000.		
XX			
PF	31-MAR-2000; 2000WO-US08621.		
XX			
PR	31-MAR-1999; 99US-0127607.		
PR	02-APR-1999; 99US-0127636.		
PR	05-APR-1999; 99US-0127728.		
PR	30-MAR-2000; 2000US-0540763.		
XX			
PA	(CUBA-) CUBAGEN CORP.		
XX			
PI	Shinkets RA, Leach M;		
XX			
DR	WPI; 2000-602362/57.		
DR	N-PSDB; AAC76229.		
XX			
PT	Novel nucleic acids and peptides derived from open reading frame X,		
PT	useful for treating e.g. cancer, proliferative disorders,		
PT	neurodegenerative disorders and cardiovascular disease -		
XX			
PS	Claim 11; Page 2727-2729; 5507pp; English.		
XX			
CC	AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,		
CC	which represent the human ORFX open reading frames 1 to 3161. The ORFX		
CC	sequences have activities such as: cytostatic; hepatotropic; vulnerary;		
CC	antipapillary; antiparkinsonian; nootropic; neuroprotective;		
CC	osteopontin; anticonvulsant; antihypertensive; immunosuppressive;		
CC	immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;		
CC	antidiabetic; hypotensive; dermatological; immunosuppressive;		
CC	antifungal; antipapillary; antiviral; antifungal; antirheumatic;		
CC	antichyroid; and antianemic. The sequences can be used for determining		
CC	the presence of or predisposition to, or preventing or treating		
CC	pathological conditions associated with an ORFX-associated disorder. The		
CC	nucleic acids can be used to express ORFX proteins in gene therapy		
CC	vectors. The proteins and nucleic acids may be used to treat cancer,		
CC	proliferative disorders, neurodegenerative disorders, osteoarthritis,		
CC	graft vs host disease, cardiovascular disease, diabetes mellitus,		
CC	hypertension, hypothyroidism, cholesterol ester storage, systemic lupus		
CC	erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,		
CC	bacterial or fungal infection, malaria, autoimmune disorders, asthma,		
CC	allergy, aplastic anaemia, burns, wounds, bone and cartilage damage,		
CC	thrombosis, contraceptive.		

CC	nocturnal haemoglobinuria, anti-inflammatory disease; to enhance
CC	coagulation; to inhibit thrombosis; and as a contraceptive.
XX	
Sequence	869 AA;
Query Match	41.5%; Score 4450; DB 21; Length 869;
Best Local Similarity	99.9%; Pred. No. 1.8e-301;
Matches	849; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Seq	Sequence 869 AA;
QY	781 RRRFLHKKAAIVFQKQKLGQIARVYRQLIARKEQKEBKKEEERERER 840
DB	20 RRRFLHKKAAIVFQKQKLGQIARVYRQLIARKEQKEBKKEEERERER 79
QY	841 ERRRAELRAQOEESTRKOQELBALQKQKAEALTRLEKQKENVQVEILRLKEIEDLQ 900
DB	80 ERRRAELRAQOEESTRKOQELBALQKQKAEALTRLEKQKENVQVEILRLKEIEDLQ 139
QY	901 RMKQOEELSTEAATLOERRDQELRLREAEACRAQAEFLSEINPBEICVNIERSL 960
DB	140 RMKQOEELSTEAATLOERRDQELRLREAEACRAQAEFLSEINPBEICVNIERSL 199
QY	961 SGSGSEFSSSELAESAACEKPNFNFSPYPREEVDGFEADDAFKDSDNPSEHSHDQRTS 1020
DB	200 SGSGSEFSSSELAESAACEKPNFNFSPYPREEVDGFEADDAFKDSDNPSEHSHDQRTS 259
QY	1021 GTRTSDSDSEEDPYMNDTVYPTSPSADSTVLAPSVODSGSLHNSGSESTYCMPPQAGD 1080
DB	260 GTRTSDSDSEEDPYMNDTVYPTSPSADSTVLAPSVODSGSLHNSGSESTYCMPPQAGD 319
QY	1081 LPSPDGYVDYDQDDYEDGALITSGSVTFNSYSQMSPDYRCASGVTNNSGAYRFSSEGA 1140
DB	320 LPSPDGYVDYDQDDYEDGALITSGSVTFNSYSQMSPDYRCASGVTNNSGAYRFSSEGA 379
QY	1141 QSSPEDEEEDFDSRFYDDELSTYRDSVSCVTLPYHSHFLYMGGLMNSKRWCVLKD 1200
DB	380 QSSPEDEEEDFDSRFYDDELSTYRDSVSCVTLPYHSHFLYMGGLMNSKRWCVLKD 439
QY	1201 ETLFMPFSKQBALKQGMHLKKGSGSSTLSRRNMKGMFVLROSKLAMYFENDSEKLGTV 1260
DB	440 ETLFMPFSKQBALKQGMHLKKGSGSSTLSRRNMKGMFVLROSKLAMYFENDSEKLGTV 499
QY	1261 EVTRAKELIINTTKENGIDIIIMADRTFHLAESPEDASQWPSVLSQVHASTDDEIOEMHD 1320
DB	500 EVTRAKELIINTTKENGIDIIIMADRTFHLAESPEDASQWPSVLSQVHASTDDEIOEMHD 559
QY	1321 EQANPQNAVGTLDVGLIDSVCASSDPDRPNSFVITANRVLHCNADTPREEMHHITTLQR 1380
DB	560 EQANPQNAVGTLDVGLIDSVCASSDPDRPNSFVITANRVLHCNADTPREEMHHITTLQR 619
QY	1381 SKGDTREGEQEFIRGWLHAEVKNPSKMSISIKLKKMFVLTHNSLDYKSSSEKNAKLGTT 1440
DB	620 SKGDTREGEQEFIRGWLHAEVKNPSKMSISIKLKKMFVLTHNSLDYKSSSEKNAKLGTT 679
QY	1441 LVNLSLCSVVPPEDEKIKETGTYNNVYTGKHCRCRLTYKLLNERTRSSVYQNTDTPKAP 1500
DB	680 LVNLSLCSVVPPEDEKIKETGTYNNVYTGKHCRCRLTYKLLNERTRSSVYQNTDTPKAP 739
QY	1501 IDPTQOOLIODIKENCNSDVVEQIYKRNPLIRLTHHPLASPLPLPYGGINLNLKDXG 1560
DB	740 IDPTQOOLIODIKENCNSDVVEQIYKRNPLIRLTHHPLASPLPLPYGGINLNLKDXG 799
QY	1561 YTTLODEAKIIFNSLQQLSMSGDPPIIQQILQTHDLRLRLDELVCQLIKQTNKVBPAG 1620
DB	800 YTTLODEAKIIFNSLQQLSMSGDPPIIQQILQTHDLRLRLDELVCQLIKQTNKVBPAG 859
QY	1621 SVGNLYSMQI 1630
DB	860 SVGNLYSMQI 869


```

XX AC ABG10632;
XX 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #10623.
XX KM Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO2001/5067-A2.
XX 11-OCT-2001.
XX PD 30-MAR-2001; 2001WO-US06631.
XX PF 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSB-) HYSEQ INC.
XX PI Dmanac RT, Liu C, Tang YT;
XX DR N-PSDB; AAS74819.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostic, forensic, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 20; SEQ ID No 40991; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridization probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. ABG00010-ABG30377 represent novel human
XX CC diagnostic amino acid sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 713 AA:
XX
XX Query Match 30.2%; Score 3238.5; DB 22; Length 713;
XX Best Local Similarity 97.9%; Pred. No. 5e-217;
XX Matches 616; Conservative 3; Mismatches 9; Indels 1; Gaps 1;
XX
XX 1337 ISVCSADSPDPSNVIIITARVULHCNADTPEMHMTLLORSQDTRVGQEFIVRG 1396
XX 1 ISVCSADSPDPSNVIIITANKVULHCNADTPEMHMTLLORSQDTRVGQEFIVRG 60
XX
XX 1397 WLHKEVNSPKKSSLLKTKRMFVLTHNSLDYKSSSEKNAIKLGTVLNLSGVVPDEKI 1456
XX 61 WLHKEVNSPKKSSLLKTKRMFVLTHNSLDYKSSSEKNAIKLGTVLNLSGVVPDEKI 120
XX
XX 1457 FKETGWNVTYGRKRCYRLYTKLNEATRWSSVIONVDTKAPIDTPTQOLIQDIKENC 1516
XX

```

```

DB 121 FKETGWNVTYGRKRCYRLYTKLNEATRWSSAIONVDTKAPIDTPTQOLIQDIKENC 180
XX
XX 1517 LNSDVVEQIKKPNILRYTHHPLSPPLPLPYGDINLMLKDKGYTLQDEAIKIFNSIQ 1576
XX
XX 181 LNSDVVEQIKKPNILRYTHHPLSPPLPLPYGDINLMLKDKGYTLQDEAIKIFNSIQ 240
XX
XX 1577 QLESMSDPIPIIOGILQGHDLRPLRDELVCQLIKQTNKVPHPGSGVNLYSWQIITCLSC 1636
XX
XX 241 QLESMSDPIPIIOGILQGHDLRPLRDELVCQLIKQTNKVPHPGSGVNLYSWQIITCLSC 300
XX
XX 1637 TFLPSRGLIKYIKFKLIRKIREPFGTEMEKALFTYBSLKKTKCEPSPSRBEIALIHR 1696
XX
XX 301 TFLPSRGLIKYIKFKLIRKIREPFGTEMEKALFTYBSLKKTKCEPSPSRBEIALIHR 360
XX
XX 1697 QEMTSVYCHGGGSGCKITINSHTTAGVEVEKIRGLAMEDSRNMPALPEYNGHVDAIES 1756
XX
XX 361 QDMTSVYCHGGGSGCKITINSHTTAGVEVEKIRGLAMEDSRNMPALPEYNGHVDAIES 420
XX
XX 1757 RTVADVLAKEFKLAATSEVGDLPWKFFYFKLYCPLDTNVPRDSVEFAPMEQAEAVIH 1816
XX
XX 421 RTVADVLAKEFKLAATSEVGDLPWKFFYFKLYCPLDTNVPRDSVEFAPMEQAEAVIH 480
XX
XX 1817 GHHPAPEENLOVLA-LKLOVYQSDYTLHAALPPLBEVYSLORLKAISQSTKTPPCR 1875
XX
XX 481 GHHPAPEENLOVLAHAATSSIMQGDYTLHAALPPLBEVYSLORLKAISQSTKTPPCR 540
XX
XX 1876 LEKRTSFLEGTLRSPRTGSVVRQVBEOKLDMWIKEEVSSARASIIDKKRKFQGMNQ 1935
XX
XX 541 LEKRTSFLEGTLRSPRTGSVVRQVBEOKLDMWIKEEVSSARASIIDKKRKFQGMNQ 600
XX
XX 1936 EQAMAKYMALIKWPGYSTLEFVCEKEG 1964
XX
XX 601 EQAMAKYMALIKWPGYSTLEFVCEKEG 629
XX
XX
XX RESULT 11
XX AAB93020
XX ID AAB93020 standard; Protein; 394 AA.
XX
XX AC AAB93020;
XX
XX DT 26-JUN-2001 (first entry)
XX
XX DE Human protein sequence SEQ ID NO:11781.
XX
XX KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX OS Homo sapiens.
XX PN EP1074617-A2.
XX
XX PD 07-FEB-2001.
XX
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,
XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX DR WPI; 2001-318749/34.
XX
XX PT Primer sets for synthesizing polynucleotides, particularly the 5602
XX PT full-length cDNAs defined in the specification, and for the detection
XX PT and/or diagnosis of the abnormality of the proteins encoded by the
XX PT full-length cDNAs.
XX

```


QY 1245 ----LMTFENDSEKLTGTEVTRAKETIIDNT-----KENGIDIIADRTPHLIAES 1293
 DB 1249 PIMLPITFMDGNTKLLADS--ATTARELCNQLSDKISLKQDFGFSLYTA--LFDKVSLSL 1304
 QY 1294 PEDDAQWFSVLQVNA-STDOEIQE-----MHDSQANP---QNAV 1329
 DB 1305 GSGGDHWDIAISQCEQYAKEQGAQENNAWRLPFRKEIAPMHEPTHDVANLIYQOAV 1364
 QY 1330 GTLDVGLDVSQASDPDRPNSFVITANRVLHCNADTFEEMHMTTLQSRKSGDTRVAG 1389
 DB 1365 RQVKG--EYRC-----DKEDLAMIYAQQYF--IEXTSMSERLFTLLPNFTIPDCLSG 1416
 QY 1390 QEFIVRGW-----LHKEVKNSPKXSLKLK-----RMFVLTNLSLDYKSEKN 1434
 DB 1417 VDKAIERNALVIAQVAKSYVWDKIAPLKIKEDIYVAKYKWPPLFSFYEAYRSGN 1476
 QY 1435 ALKLTGLVINSLS--VPPDEKIFKETGYWNTVYGRHCHRLYKLNENT----- 1485
 DB 1477 LFKNDVILAVNWTGVYVDDQOVLLLESPPEITAVSSQKTKNVFTQPSLSLTVRGEPT 1536
 QY 1486 -----RMSVIOQVTDTPKAPIDPT-----QOLIDIKENCIN 1518
 DB 1537 FQSPNABEDRDVAVYFPLDGLKKRSKYVIALODYRABSDTSTFLSPFKDILLDESCGE 1596
 QY 1519 SDV-----VEQIYKBNPILRYTHPLHSPLLPLPYGDI-----NLNLK 1557
 DB 1597 SVLNNQMCIGRCDBRSGERDPAETVYVL--PLSKRPQDILAFNIEBAHNGRLSNMS 1654
 QY 1558 DKGYT-----TLQDEAIKINSLOQLESMS-----DPI--PIIO- 1589
 DB 1655 NGAVEPRDRPHTLMEYALDHR--LPPKRTMSKTLTSLSKRSBELWRSRDIKAPLARK 1713
 QY 1590 -----GILQTHDL--RP-----LRDELVCOLK 1611
 DB 1714 LOSKEPFAEACPAFAAILKTMGDPLSKPRMGNETTDIIFQGPLKHELRLBEIYQOLK 1773
 QY 1612 QTNKVPHPGSGVNLXSWOILTCLSTPLRSGLKYLKYLKRIROQFTEMEKALFT 1671
 DB 1774 QL--TDNRMRSEBERGMELMWLATGLFACSGGLKELFLKTRRRRPIQDSMHR----- 1826
 QY 1672 YSLKKT--KREBPVSRDEIEALIR--QEMTSTYCYGSGGSCKTINSHTTAGEVAK 1727
 DB 1827 ---LQKTIHQGRKYPHQVEVEALIOKTOIFHKYFPDDTDEAFEVDSSTRAKDFCNN 1883
 QY 1728 LIRGLAMEDSRMFPAL-----FEYNHGVDAIESRTVAVAVLKFEKL 1770
 DB 1884 ISQRLSRTSEG--FSLFVKIADKVISVPRGDFPDRHLTWIKKARPIRD----- 1934
 QY 1771 NMTSEVGDLPKRFYFLYCF--LDTDNV--KD--SVEFAFMFEQAHAVIHHHPAPENTL 1826
 DB 1935 -----GANP--QFTYGVFFMKKLMNTVPGKORNDLIETHYQELPKLGRYHKCSREBA 1987
 QY 1827 QVLAALRILOYLOGDYTLHAIRPLEEVYSLQRIKARIESGTTFTPCERLEKRTSFFLEG 1886
 DB 1988 AKLAALVFVRPGE-----NKOELQA--IFQMLRELIAPSDIMKIOSTS----- 2028
 QY 1887 TLRBRPRGVSVRQKVEEOMLDMWIKESVSSABASIIIDMKRFGQMNQEQAMAKYMALI 1946
 DB 2029 -----EW-----KSTIVASYNQDGMSTEDKVAALVK 2057
 QY 1947 KEMPGYSTLPDV--ECKEGGFPOELMIGVADAVSYKAGEERPLEVFQYEHILSPGAPL 2005
 DB 2058 YMPFPGSAFPEVKQTEBPYEMMLIAINKGVSIIHVTQDILVTHFTTISNMS-- 2115
 QY 2006 ANTY-----KIYVDERELLFETSEVVVDVAKMKAYSNIV 2040
 DB 2116 GNTYFMTIGNLVRGSKLCESTISLYKMDLLTYSIAML 2155

AC AAUI7508;
 DT 07-NOV-2001 (first entry)
 XX
 DE Novel signal transduction pathway protein, Seg ID 1073.
 XX
 KW Neutropenic; cytostatic; dermatological; immunosuppressive; tumour;
 KW antineoplastic; anti-ILV; antibacterial; antiinflammatory; cancer;
 KW immune system disorder; rheumatoid arthritis; inflammatory condition;
 KW organ transplant rejection; infection; hepatitis C; blood disorder;
 KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
 KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;
 KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
 KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
 KW reproductive system; gastrointestinal; liver disorder; AIDS;
 KW acquired immune deficiency syndrome.
 OS Homo sapiens.
 XX
 PN WO200154733-A1.
 XX
 PD 02-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01312.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 11-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 14-AUG-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225266.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0228927.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229345.
 PR 05-SEP-2000; 2000US-0229503.
 PR 06-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 08-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.

RESULT 14
 AAUI7508
 ID AAUI7508 standard; Protein; 348 AA.
 XX

PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246529.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.

PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250161.
PR 05-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251857.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-465460/50.
DR N-PDB; AAS27425.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT diagnosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX
XX Claim 1, SEQ ID No 1073; 880bp; English.
XX
XX The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX
Query Match 16.7%; Score 1789; DB 22; Length 348;
Best Local Similarity 99.1%; Pred. No. 2,5e-116;
Matches 335; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1033 DPYNDTVTPSPASDSTVTLAPSVQDSSGLHNSGSGSTYCPMPNAGPLSPDGDYVD 1091
DB 6 DAWVNDTVTPSPASDSTVTLAPSVQDSSGLHNSGSGSTYCPMPNAGPLSPDGDYVD 65
QY 1092 QDDYEDGAIITSGSSVTFSNSYSGQWSPYRCGCVGTYNSSGAYRFSSEGNQSSFFEDSEDF 1151
DB 66 QDDYEDGAIITSGSSVTFSNSYSGQWSPYRCGCVGTYNSSGAYRFSSEGNQSSFFEDSEDF 125
QY 1152 DSRPDTDELSTYRDSVYSCTVLPYFHSFLYWKGGGLNWSKRRKCVLKDFTFLMRSKOE 1211
DB 126 DSRPDTDELSTYRDSVYSCTVLPYFHSFLYWKGGGLNWSKRRKCVLKDFTFLMRSKOE 185
QY 1212 ALKQGMHLKKGSGSTLSRRNKKRPFVLRQSKLMPFENSEKELKGTVEVTAKETIDN 1271
DB 186 ALKQGMHLKKGSGSTLSRRNKKRPFVLRQSKLMPFENSEKELKGTVEVTAKETIDN 245

QY 1272 TTKENGIDIMADRPHLLAESPEDASQWFSVLSQVHASTDIOIEMDEQANPONAVCT 1331
 DB 246 TTKENGIDIMADRPHLLAESPEDASQWFSVLSQVHASTDIOIEMDEQANPONAVCT 305
 QY 1332 LDVGLIDSCASDSDPSPNSPVITANRVLHCNADTPE 1369
 DB 306 LDVGLIDSCASDSDPSPNSPVITANRVLHCNADTPE 343
 RESULT 15
 ABB62828
 ID ABB62828 standard; Protein; 2129 AA.
 AC ABB62828;
 DT 26-MAR-2002 (first entry)
 DE Drosophila melanogaster polypeptide SEQ ID NO 11276.
 KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.
 OS Drosophila melanogaster.
 PN WO200171042-A2.
 PD 27-SEP-2001.
 PP 23-MAR-2001; 2001WO-US09231.
 PR 23-MAR-2000; 2000US-191637P.
 PX 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 PI Venter JC, Adams M, Li PWD, Myers EW;
 DR WPI, 2001-656860/75.
 DR N-PSDB; ABL06931.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 PS Disclosure; SEQ ID NO 15276; 21bp + Sequence Listing; English.
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
 CC sequences (AB101840-AB16175) and the encoded proteins
 CC (AB57737-AB12072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 CC XX
 SQ Sequence 2129 AA;
 Query Match 15.9%; Score 1705; DB 22; Length 2129;
 Best Local Similarity 25.7%; Pred. No. 3.1e-109;
 Matches 598; Conservative 392; Mismatches 828; Indels 508; Gaps 84;
 QY 3 LOGTAVLRENGQHPESTVNSCAE-----GIVFRFDYGVFTR---YQSTITTHQKV-T 52
 DB 13 IQGSEYVWVWPQ-----NTTSEFVAVPFGARIVREKIQTLVCDNRKQFWMPADVIK 64
 QY 53 AAHPTEEGVDVMASTELHGSIMYNLFOYKRNQIMTYGISILASVNPYQPIAGLYEP 112
 DB 65 AHHITSGEVEDMTITGDLQEVITILNLONRVAKQILYITTSMLVAIWPQ-ILPIYTN 123
 QY 113 ATMEQYSRRHIGELPPIFAIANECYRCIMKGRHNDQILIKESGAGKTESKILIKFLS 172

DB 124 REIQLYRKSJAELEPPHIFAISDNAFORLQRIKENQCVVISGESBAGTESKILIQLYA 183
 QY 173 VISOQSLELSKEKTSQCYERAIILESSPIMEAFGNKATVYNNNSRFGKFOVNIQOKNI 232
 DB 184 AIS-----GKSWIEQOIIENAPIMEAFGNKATVNDSSRFGKTEIRFPQGI 234
 QY 233 QGGRIVDILSQNRNVNQRNGEENYHIFVALAGLEHEERBEFYLL-STPENHYLNQS 290
 DB 235 QGARIQOYL- EKSLIVFQSRDERNYHIFCYCLAGLSTAHERRLTQSQSPQHYHLLQ 293
 QY 291 GCVEDKTSIDQSEFREVITAMDVMOFSKEEYREVSRLLAGILHGNIEFITA----GGA 345
 DB 294 GCFLLPGRGDAKDFADIRAAKVLSPKPEEWSILSLAAILHGLNLTREFTAEVANLATA 353
 QY 346 QVSEFTALGRSAELGLPQTQTLDTLQSRSMPLREELTLPVNOQVADSRLMALAYA 405
 DB 354 EIDTTPNLRVAQQLGIFISALNALVORTTFVGEHVITSLSKNALEGRAFYKSLYD 413
 QY 406 CCFEWWIKINSRIKGNED--FKSIGIIDIFGFENFEVNHFEQFNINYANEKLAQEFYFNH 463
 DB 414 GIFVAVIRINETINKQVDQPNNSIGVLDIFGFENFDNNSPFOQLCINANENLQGFVGH 473
 QY 464 IFSLEQLYSRGLVWEDIDWIDGECDLI-EKKGLALINBSHPQATDSTLLEKL 522
 DB 474 IFKMEQDEYONHIMQHIEFODNOQIIDLIGMKPMNLSLIDSESKPEKGTDTLTLEKL 533
 QY 523 HSOHANNHFFYKPERAVANN-FGVKHAYGEVQVDVAGILEKRDPTFDLLNLTRESRPF 581
 DB 534 HVQGNRSITVYKAGTKTQTSLEIRHYAVGVVNNPLGFEKNDSDSGDRTLVQSTNKY 593
 QY 582 IYDLFEHVSRRNQDTLKGSGKRRPPTVSQFQVDSLMLATSSSNPFPRCKPKPMQ 641
 DB 594 LVDIFPH---EMPQDTAK-----KQPLTCVFR-NSLMLWRTLSQAHPIFRCKPKPEY 644
 QY 642 KMPDQDAVAVLNQIRYSGMLETVAIRKAGVAVRRPFQDFYKRYVLMKRLALPE--DVR 699
 DB 645 KEPKPFDELQVRQLRYSGMETAIIRAGYPIRHAIVRAFERVRLVLPVGPQLQCOCR 704
 QY 700 GKCTSLQLDYASNSEMOGLKTVFLRESLQKLEKEEBSHAAVIRAVLGLFLARK 759
 DB 705 KLAQICVAVLPADSDRGYKTKFLRDEDDASLELQSLQSLKSIIVTIQRIIRVLFPR 764
 QY 760 QYKRVLYCVIILQKNYRAFLLRRLRLHLK-----AAVLPQOL----- 798
 DB 765 YMKRREALITVQRYWRGLQRRKQVWRQGNHRLGALIAQQLITKXTMWRCTITXQA 824
 QY 799 --RGQIARVYRQLLAERKEQEEKKQD-----EERKKQEEERERER 840
 DB 825 LSRGYLVKRDFOKLLERRKQQLKKEILLKAKWKAEBELLRLQOLKEQKEREQEOE 884
 QY 841 EREKELEAQQEERFKQOELEAOKSKEAELTRELK-----QKEN----- 883
 DB 885 KRLQEEQLKKAARNAALMAAAVQAKRRTYVQKQAPLTLQNRSLPPPTTLIVAA 944
 QY 884 ----KQVEEILRL-----EKIEDLQR--MKEQQELS-----LTEAST-- 915
 DB 945 PLPTPPAAVTRINTIPSPGCTIVESSKQWVDVFRFLNDEPDSGLPGPVKSKMMP 1004
 QY 916 -QKQERDDELRLREBEACRAAOEFLSLNFEIDECVR-----NIERS---LSGG 963
 DB 1005 EQALRLRDRDVPFKLLSRPAA-----LRKLNISSGDTRILPKSVNNIDTDSFYLKVA 1058
 QY 964 SEFSSELEASACEKPNFNFGQPYEEBVDSEFADDDAFDSPIPSHGHSDQTSIGI- 1022
 DB 1059 ATYFGGATLQHERP-----LKKSLAKHHPIDEMASKAIW 1095
 QY 1023 ----RTSDSSSEEDPYNDTV-VPTSPSADSTVLLAPSVQSGSLHNS--SGESTYCMQ 1076
 DB 1096 LTIIRFMD-----LDDVSSPTLHVFNDENLMS-----DLASLNTSDSKRPLFRQ 1144
 QY 1077 NAGDLPSEGDVDYDQDDYEDGALITSGS-----VTFNSYSG-----SQWSPDY 1120

1145 SQRRIKPLASGEKEAEFYQHMLNVPSTLSEKIHFIIGHIINKSLRCFPSPKEFEPHL 1204
QY 1121 RCGVGTYNSSGAVRFSSEG-AOSSFEDSEEDPDSRDTDELSYRDSVSCUTLYFYS 1179
DB 1205 R-----SFMKQGTALQATPSLQRLERTLVNGPRC--QPPSLBELHAINGRHP 1250
QY 1180 FLYMKGLMNSWRKRCVLDKDETFELWFRSKOEBALKO-----GWLH 1219
DB 1251 -LRDIDLMDGQGRRLQVDAAST-----AREAVNLCOQMGMLDTFGGLVMSLNGKLM 1303
QY 1220 KKGGS-----STLSRRMKRW--FVLROSKLMTFENDSEK-----LK- 1257
DB 1304 PLGAGEHVLDAISEBORQDAPMKLYRKEMFATWYDPSMDPKATOLYKQILNGLKC 1363
QY 1258 GTVEVTRAKEIIDNT-----KENGIDIIMADRTFHLIASPED-----ASQWFSVLS 1305
DB 1364 GEYRCSEKDIAWVCALACVEYGPGEILKPKSEITAFVPSDLLAPGERAIENMSRLIA 1423
QY 1306 QVHASTDQEIQENHDEQANPONAVGTLIDVLCASDSPDRP---NSFVIITANRYL 1361
DB 1424 ATYEKSSVYKEQNDLLLEAQKRAK-----EDICLFAHLSWPMRHSRLFVV----- 1470
QY 1362 HCNADTPPEMHMTITLQSKSDTRVEGGEFIVRGWLHKE-----VKNSPKSSLKTKR 1416
DB 1471 --RKEGPKLQSDLEMLGINSAGFLIDEFQVILASCCFSEVULKVHVESDDKLVMTFQHV 1528
QY 1417 WFLYTHNS-----LDYKSSSEKALKGLTVLNSL--CSVPPDEKIFKETGYW 1463
DB 1529 NFVLQCSSADANEVINYMLDNIKORSYGVALDPVVEGDLEDCLVANFGDLIEFAGVT 1588
QY 1464 NVTYV--GRKRCYRLTKLNEATRWSVYQNTDTKAPIDPTQOLIODIKENCUN--- 1518
DB 1589 GAQLMAGNADCYR---GCYN--GQWQFLAGNVRVLAITLKPSKLODILREGRFQBP 1643
QY 1519 -----SDVVEQIYK-----RNPILRYTHPLASPLPLPYGDINLNL 1556
DB 1644 KPTPRANYSRRQHNISQLAESHFRPLSDKAPLSKFSPEPLKAPL----- 1691
QY 1557 KDKGYTTLDEBAIKINFNSLOQ-----LESMSD-----PIP---IIQGIQTHDRL 1599
DB 1692 -----KAVVKVPPLFOQALVMHHIILIKMGDIARSNLPVNTDLIFOPALO--HPL- 1739
QY 1600 PLDELYCOLIKQTNKVPHPGVSQNLYSQILTCLSTFLPS---RGILKYLKFN--- 1651
DB 1740 -LCDELYCQMKQLS--DNPSSESEKRGMDLVLATGLVAPSVLVMRELIILMRADAL 1796
QY 1652 ----LKRIREQPGTEMEKA-LFTYESLKTKC-----REFVPSRDEIBALIHROEMTS 1701
DB 1797 ADACLRLKRSLSAQGRKAPHLIEVEGIQO--RCLHIYHKIYFPD-DTVEA----- 1845
QY 1702 TVYCHGGGSCKITINSHTTAGVEVEKLIINGLAMEDSRNMFALPEYNGHVDKAIESRTVA 1761
DB 1846 -----FEIESHTGAELIADIQAQLELK-SPVGYSIFLKTGDRYAMPEEEFVF 1893
QY 1762 DVLAK---FEKLAATSEVGDLPWKFFYFKLYCFCLDTDNVPKOSV--EFAFMEOAHEAVI 1815
DB 1894 DFITQIYMLRQORTIRSIDGQYQHFMRKML--NNHGEDLNGDMIFSYQBELHXYL 1951
QY 1816 HGHHPAREENLOYLALRYLOGDYTLHAATPPLLEVYSLORLKARISOSTTFTPCER 1875
DB 1952 KGYYPIDCQASRLAL-----VYSADHDVSLQRLP----- 1982
QY 1876 LERKRTSPLEGLTARSFRGVSVRKQVBEQMDMWIKEEVSARASIIDKMRKFGQMNQ 1935
DB 1983 -----EVLTRILPE---DLIPLOTVAEWRQOILPKVHR-DHLYE 2017
QY 1936 EQAMATYMLLIKEMPGYSTLPDV--ECKEGFPQELMLGVASADAVSVYKRGGRPLEVFO 1994
DB 2018 DHAKIIFLOELSHFACGSGTFVVKQONDALPETLIINSTGFMHLDPTTKEILRSYE 2077
QY 1995 YEHILSFAGAPLANTYKI---VDERELLFETSEVVDVAKMKAYI 2036
DB 2078 YSO-LGIWSSGKNHFHIRGNGMIGASKLCTTQGYKMDLLASYV 2122

Search completed: July 14, 2003, 18:15:22
Job time : 80.3175 secs